THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 46, PART 1 *January* 1, 1934 NO. 252

538.15

THE MEASUREMENT OF MAGNETIC FIELD DISTORTION

Presidential Address delivered on November 17, 1933 *by*

PROFESSOR A. O. RANKINE, O.B.E., D.SC., F.INST.P.

ALTHOUGH it has not been the custom in recent years for Presidents to make particular reference in their addresses to the affairs of the Society, I feel that I must on this occasion depart from precedent to mention prominently one event of exceptional importance which has occurred during my period of office—I mean the union with the Optical Society. Of the wisdom of this step there can, I think, be no question. It was taken only after the Councils of the two Societies had assured themselves unmistakably of the mutual desire for amalgamation on the part of the members. The formal union took place in October 1932, and, at the end of March of the present year, the bulk of the members of the Optical Society had been received into the Physical Society—the modified name adopted by the Physical Society of London. Our Society now numbers more than a thousand members, and is, I am sure, much stronger and more active and useful than at any time in its previous history. As an officer continuously for fifteen years I have watched this growth from within with mixed feelings of satisfaction and pride, and to have the honour of being President at such a time is particularly gratifying to me. I believe our Society has before it a still greater future, in which I am still young enough to expect to share.

I have another reason for referring to the union with the Optical Society, namely, that I was President of that Society also. Luckily for me that Society ceased to have a separate existence in time to let me escape the duty of delivering a presidential address. Otherwise I should have had three addresses to prepare in the space of little more than a year, having presided also over Section A of the British Association in 1932. This escape left me with the half-formed intention to rectify the matter by choosing for my present address an optical subject. Indeed, I think I mentioned this privately to a number of Fellows. I have now to announce the abandonment of this project, and plead in excuse inability to find the necessary

time for preparation, optics being no longer the work upon which I am engaged. For several years I have been particularly interested in the relatively new subject of applied geophysics, and since 1931 I have been occupied in trying to establish in the Imperial College a small department for study and research in this subject. Accordingly I have chosen the easier path of addressing you now upon certain aspects of the work with which I am immediately concerned in the course of my ordinary duties.

I propose to discuss the possibility of improving the methods at present in use in magnetic surveying in relation to the determination of hidden geological structure. It is a subject definitely terrestrial, even trivial compared with the universe discussed by my brilliant predecessor. To be called upon to follow him weighs on me very heavily, and you must forgive me if the manner of my address merits the same comparison as the matter.

In the course of preparing lectures in the new department I have mentioned, I was brought intimately into touch with the torsion balance invented by Baron von Eötvös for measuring gravitational anomalies. Like all, I suppose, with similar experience, I became filled with surprise and admiration at the beauty of principle and practical efficiency of instruments of this type, and was able to present to the Society last year two papers on certain new aspects of theory and practice*. As a means of determining with precision the small variations of the earth's gravitational field the Eötvös torsion balance has become thoroughly established and is widely in use. I am indebted to Mr Lancaster Jones for the suggestion, which is at the root of what I have to say, that it might be possible to construct a *magnetic* torsion balance embodying the virtues of the gravity instrument, and thus enabling local magnetic variations to be measured in a similar manner. An examination of the literature revealed the fact that Eötvös himself had a similar idea—indeed that he actually constructed such a balance†. He does not appear, however, to have proceeded far with it, probably on account of certain practical difficulties. Later on I shall have to lay stress on these and admit that so far my experiments have not shown the way to eliminate them. It may even be that the object will not be attained until someone more capable turns his attention to the task. Yet for that very reason it seems worth while to formulate the great advantages which would attend a successful issue. Moreover, there does emerge from what has already been done the promise of a partial achievement not without interest in connection with the measurement of magnetic susceptibility.

There are, of course, already in use a number of portable magnetic variometers designed to indicate small local magnetic anomalies in particular regions of the earth's surface. Of these may be mentioned as typical the quartz-fibre horizontal-intensity variometer‡ used by Walker in his magnetic survey of Great Britain, the gravity-compensated horizontal and vertical magnetic variometers, due to Schmidt and now in very common use in geophysical surveying, and the recent adaptation to portability of F. E. Smith's current-compensated instrument for measuring the horizontal intensity. In all such instruments the small variations of magnetic force

* *Proc. Phys. Soc.* **44**, 465–489 (1932). † *Ann. der Phys.* **59**, 373 (1896).

‡ This was indeed a torsion balance, but not depending on the principles here contemplated.

in passing from one station to a neighbouring one are made evident by balancing nearly the whole of the earth's magnetic control through the agency of a compensation subject to artificial adjustment. Thus in the Walker instrument the compensation is effected by applying torsion to the suspending quartz fibre; in the Schmidt variometers the magnetic couples are balanced by suitably disposed weights, and an opposing magnetic field procured by passing an electric current through a pair of coils arranged in the Helmholtz manner provides the necessary nearly complete compensation in the Smith apparatus. By means of these methods it has been possible to construct instruments capable of detecting under favourable conditions field-differences of rather less than 10^{-4} *gauss** or 10γ at any rate in the horizontal intensity. But it is important to note that this degree of sensitivity is only attained as a result of nearly complete compensation of the main field.

A serious limitation is imposed by the effects of variations of temperature on the readings of these variometers. In geophysical practice wide changes of temperature have to be faced, and it is important to escape, if possible, from spurious effects arising in this way. Consider, by way of illustration, what is the effect of temperature upon a horizontal variometer of the Walker type. This instrument consists of a magnet suspended with its axis horizontal by means of a quartz fibre upon which there is imposed by means of a torsion head sufficient twist to hold the magnet approximately at right angles to the meridian. If the temperature changes, both the torsional coefficient of the fibre and the moment of the magnet are altered, and the equilibrium is disturbed to an extent proportional to the full value of the earth's horizontal field—a disturbance which may be much greater than the small genuine effects, the measurement of which is contemplated. In the Schmidt instruments temperature similarly affects the linear dimensions of the balancing arm of the compensating weight as well as the magnetic moment; and in the Smith apparatus there are variations of dimensions of the Helmholtz coils and of the e.m.f. of the standard cell used for current measurement. It is true that to some extent appropriate corrections can be applied, and that, indeed, in the Schmidt instruments, an ingenious temperature-compensating device has been incorporated. But it would clearly be advantageous to reduce the actual temperature effect to second-order proportions so that it would never be more than a small fraction of the small quantity measured. In the magnetic analogue of the Eötvös gravity balance this would be achieved.

A still more important advantage would accrue from the realization of a magnetic torsion balance corresponding to the Eötvös gravity instrument, namely, that it would differentiate between spatial and temporal variations of the earth's magnetic field if the latter were sufficiently wide-spread; it could deal with just such fluctuations as are experienced daily on the earth, and even magnetic storms. It would respond fully to the variations of field from point to point, but inappreciably to general time variations of the same order. Although the question does not really arise on the gravitational side, it is interesting to contemplate that if the earth's

* I am not sure whether it is still proper to call the field unit the *gauss*; I believe it should now be the *oersted*.

gravitational intensity were subject generally to small changes with time, the work initiated by Eötvös would not be handicapped. As it is, his instrument has not had the opportunity of displaying one of its chief virtues. In the magnetic case the position in this respect is, of course, different. It is a feature of the earth's magnetic field that it undergoes appreciable diurnal variations both in direction and magnitude, and occasionally there occur magnetic storms of still larger extent. As a result there are superimposed upon the magnetic effects of buried structure disturbances by no means negligible, which affect all ordinary magnetic variometers and mask the true anomalies sought. No artificial compensation is here possible, because the extent of the time variation is not known beforehand. The result is that it becomes necessary to apply corrections of rather uncertain amount by subsequent reference either to an established magnetic observatory or to an additional instrument set up at a fixed station for the purpose. This complication would vanish if we could use instead a magnetic torsion balance truly analogous to the gravity instrument, and at the same time the sensitivity of magnetic measurement would be greatly enhanced.

These important advantages of the Eötvös torsion balance—its very great sensitivity and its nearly complete independence of time variations of temperature and of field—are due to the peculiar nature of the compensation employed. The instrument is, in fact, self-compensating. In a spatially uniform field it adjusts itself automatically to a quiescent state, there being under these conditions no torsion in the suspending fibre. The line joining the point of support to the centre of gravity of the suspended beam becomes naturally a vertical axis, and the only forces which are operative in twisting the fibre are the small horizontal components of the gravitational attractions on masses displaced from the axis. Advantage is taken of the great sensitivity attainable in fibre suspension, and only the variation of the field—the main field itself being excluded—affects the orientation of the beam. The only limitation on sensitivity which gravity itself imposes is that it necessitates a fibre or wire strong enough to support the beam, and thus prevents indefinite reduction of the torsional coefficient.

I want to explain now how, at least theoretically, the principle of the gravity balance could be applied to an equivalent magnetic instrument. To do so may I remind you first of what the gravity balance actually measures directly. If it is simply a Cavendish balance, for example a suspended horizontal beam bearing at its two ends two equal masses, it measures those small distortions of the gravitational field represented by the quantities $\partial^2 U/\partial x^2 - \partial^2 U/\partial y^2$ and $\partial^2 U/\partial x \partial y$, where U is the gravitational potential and x and y are two arbitrarily chosen horizontal rectangular coordinates. If the beam system is modified, as in the Eötvös instrument, by displacing one of the masses vertically with respect to the other, e.g. by suspending it at a lower level from one end of the beam, two additional quantities are measured, $\partial^2 U/\partial x \partial z$ and $\partial^2 U/\partial y \partial z$, z being the vertical direction at the centre of gravity of the beam. Since $\partial U/\partial z$ is the local value of g, these two departures from uniformity of field are clearly $\partial g/\partial x$ and $\partial g/\partial y$, which we identify with the horizontal components of the gradient of gravity. We can, if we desire, express these in the form of a single

U

x, y

z

g

vector in the direction s making an angle ψ with x by the relations s, ψ

$$\frac{\partial g}{\partial s} = \sqrt{\left\{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2\right\}} \text{ and } \tan\psi = \frac{\partial g}{\partial y}\Big/\frac{\partial g}{\partial x}.$$

For some purposes, as I have shown elsewhere*, it is indeed convenient to regard $\partial g/\partial s$ and ψ as primary magnitudes derived from the instrumental observations.

Now if we can copy the gravity balance so as to obtain a magnetic equivalent, corresponding magnetic quantities will be measured since U will now denote the magnetic potential instead of the gravitational. For example, the gravity-gradient will become the gradient in a horizontal direction of the vertical component of the local magnetic field. To impart to the balance these new functions it is necessary to substitute magnetic poles for gravitating masses, with due regard of course to the polar nature of magnetism—a property not displayed by gravitation. A consideration of the means of doing this shows that it would be preferable not to try to construct a single magnetic balance fully equivalent to the usual form of the Eötvös instrument, but to aim at making two separate instruments, one to measure the magnetic quantities $\partial^2U/\partial x^2 - \partial^2U/\partial y^2$ and $\partial^2U/\partial x\partial y$, and the other to measure the magnitude and direction of the horizontal gradient of vertical magnetic intensity. A further reason for this division is that the second form seems to be more capable of practical realization, and may be conveniently considered by itself.

The magnetic analogue of the Cavendish balance would consist ideally of two equal magnetic poles of the same kind (say N) situated at or near the two ends of the horizontally suspended beam. Owing to the impossibility of isolating single poles whole magnets must be used, and the unwanted poles of the opposite sign must be so disposed as to produce relatively little turning effect. Two equal magnets, for example, with their south poles adjacent and near the axis of suspension, might be arranged with their axes directed precisely in opposite directions, and so form the beam. In a uniform field there would be no resultant turning moment in any azimuth; but particular kinds of spatial non-uniformity would give rise to a differential couple tending to twist the suspension, just as in the gravitational case. I have examined in some detail the magnitude of the small forces operative, in order to compare the possible sensitivity of the arrangement with that of the corresponding gravity balance. That is to say, account has been taken of the attainable pole-strengths in relation to the weights of the magnets possessing these poles, and the supporting strengths and torsional coefficients of available suspending fibres. I find that, with the very high values of remanent magnetism which can be possessed by cobalt steel, very great sensitivity can be anticipated. The relative position as regards magnetic and gravitational fields may be stated in this way. If we regard unit magnetic pole as equivalent to unit mass in the c.g.s. system, then, apart from the peculiar difficulties of construction or limitations of application to be mentioned later, the magnetic balance could be made several times more sensitive than the gravity balance. And, just as Eötvös found it convenient to define a small unit, 10^{-9} dyne/gm.cm. called the Eötvös unit, to express the small measured values of

* *Loc. cit.* and Presidential Address Section A, British Association, 1932.

such quantities as $\partial^2 U/\partial x^2$, we should need the corresponding magnetic unit 10^{-9} gauss/cm. for practice with a magnetic balance, if eventually forthcoming.

An interesting question arises in this connection. In field practice with the Eötvös gravity balance, corrections always have to be made on account of gravitational non-uniformity arising mainly from the rotation of the earth, and not attributable to local density-structure. One of these involves the quantity $\partial^2 U/\partial x^2 - \partial^2 U/\partial y^2$ which, except near the geographical poles, has an appreciable value. Physically this implies that the Cavendish beam system will tend to set perpendicular to the meridian of longitude. We may ask if there is any corresponding tendency in the magnetic case, in so far as we may regard the earth as a uniformly magnetized sphere. Calculation shows the answer to be in the negative, for both the quantities $\partial^2 U/\partial x^2 - \partial^2 U/\partial y^2$ and $\partial^2 U/\partial x \partial y$ prove to be zero under these assumptions; but I have to confess that so far I have been unable to picture physically why the astatic system described earlier should be in neutral equilibrium everywhere on the earth's surface. The result, however, is of significance. It means that the only deflections observable on a magnetic balance of this type would be those due to field-distortion arising from the local magnetic deposits which are the objects of search.

But I must proceed no further with this possibly too enthusiastic description without exposing the fly in the ointment. There is a very serious practical difficulty of construction which is quite likely to prove insuperable in respect of producing a field instrument for use at any arbitrarily chosen place. It arises from the fact that the earth's magnetic field is not everywhere nearly vertical. However useful the horizontal component may be to mariners, it is an unspeakable nuisance when we want to make really delicate magnetic measurements. Nor can the usefulness of the tangent galvanometer for instructional purposes reconcile me to H, whether properly specified in gauss or in oersteds. The existence in ordinary latitudes of this large horizontal magnetic force makes demands on our capacity for fine adjustment which seem likely to rob us of a large part of the sensitivity otherwise attainable, as well as making impracticable the use of magnet torsion balances following the normal procedure with gravity instruments. I can illustrate this by giving brief particulars of an instrument which I proposed at one time to construct. It comprised a metallic strip suspension 20 cm. long having a torsional coefficient of 0·007 dyne-cm. per radian. This had a breaking strength of 55 gm. wt. and was designed to support in the manner already indicated two cobalt-steel magnets each 5 cm. long and of mass 10 gm. Such magnets can have in practice moments of approximately 600 c.g.s. units, and, ideally, the completely astatic system contemplated would have displayed a magnetic sensitivity for indicating field-distortion of approximately 2 scale-divisions per 10^{-9} gauss/cm., as compared with $\frac{1}{10}$ scale-division per 10^{-9} dyne/gm.cm. in the original Eötvös gravity balance having the same optical magnification. In other words, for equivalent magnetic and gravitational field non-uniformity, the sensitivity would have been multiplied 20 times.

Only the difficulty of attaining the degree of astaticism to render the suspended system irresponsive to a uniform horizontal field prevents this precision being realized. But it is one so serious as to be, I fear, insuperable. In order to reduce the

effect of a uniform field of, say, 0·18 gauss to negligible proportions, i.e. to 1/10 scale-division, it would be necessary for the two magnets to have moments equal, and persisting in equality, to an accuracy of about 3 parts in a thousand million. Moreover, the alignment in opposite directions of the two axes must have a precision of less than one-thousandth of a second of arc. Of course, with smaller values of H the demand becomes less stringent. Indeed, in the immediate neighbourhood of the earth's magnetic poles the system could possibly be used. But these regions are limited—perhaps of radius a dozen metres—and we have no reason to suppose that geological deposits of great economic importance lie in these isolated parts. The whole argument, however, although disappointing in its results, is a remarkable illustration of the delicacy of the torsional suspensions which we have at our disposal.

I pass now to the consideration of another system which is not so unpromising —what may be called by analogy a magnetic gradiometer. In gravity surveying we have the gradiometer of Shaw and Lancaster Jones* in which the beam is designed so as to exclude what the Cavendish system measures, and to respond only to the gravity-gradient $\partial g/\partial s$. This instrument is sensitive enough to measure the gradient of gravity in a northerly (or southerly) direction, arising mainly from the rotation of the earth, which has a maximum value of about 8×10^{-9} dyne/gm.cm. in the latitudes 45° and is zero at the poles and the equator. The corresponding quantity in magnetism is $\partial V/\partial s$, V being the vertical component of the earth's magnetic field. V
Taking the earth as uniformly magnetized, calculation shows that $\partial V/\partial s$ varies from the maximum value of about $1{\cdot}5 \times 10^{-9}$ gauss/cm. at the magnetic equator to zero at the magnetic poles. The design of a magnetic gradiometer capable of measuring this gradient easily presents no difficulty save one, again due to the bugbear H, but not, I think, so great, which will be examined later. Indeed, the design in this case becomes particularly simple as compared with that required for gravity. Imagine a magnet hung with its axis nearly vertical from one end of a non-magnetic beam supported at the centre by the usual suspension and counterpoised with a non-magnetic mass attached to the other end. In a uniform magnetic field the resultant translational force on the magnet would be zero, since the poles are precisely equal and opposite in strength. It could only have a horizontal component in so far as H varied in a vertical direction, and it would thus be proportional to $\partial H/\partial z$ or $\partial^2 U/\partial s \partial z$, where s is the horizontal direction of maximum variation of H s
and z is vertical. The moment of this force would twist the central suspension to an z
extent depending upon the azimuth of the beam; thus the deflections would provide a measure of $\partial H/\partial z$. This differential coefficient can, of course, be identified with $\partial V/\partial s$, since both are equal to $\partial^2 U/\partial s \partial z$; accordingly the measurement of the horizontal gradient of the vertical magnetic intensity is implied.

One interesting point of difference between this system and the gravity analogue may be noted. In the latter, there being only one kind of mass, the two effective masses have to be arranged at different levels on *opposite* sides of the central suspension in order to isolate the differential effect to be measured. (This happens to

* This has recently been described fully in the *Journal of Scientific Instruments*, **9**, 341 (1932).

be convenient also for counterpoising purposes.) The polar nature of magnetism requires for the corresponding effect that the two opposite poles should be on the *same* side of the suspension as indicated; luckily therefore no matching of different magnets is required as in the Cavendish form of instrument, and counterpoising can be effected by relatively non-magnetic materials. While on this point I may mention, in order to dispose of it, the question of possible confusion of magnetic and gravity effects. We cannot have magnets without inertia, and, in general, every magnetic torsion balance will be responsive also to gravity effects. Calculation shows, however, that, having regard to permanent magnetism in relation to mass in cobalt steel, it would be easy to secure that the gravity response shall be only a few per cent of the magnetic. Moreover, if necessary, advantage could be taken of the distinction between magnetism and mass by arranging such a distribution of the two in the beam as to render the system practically inert to gravitational anomalies.

If I give some details of a magnetic gradiometer I have designed and begun to construct, it may serve to illustrate further discussion on various points. The beam is 4 cm. long and has suspended from its ends by fibres of lengths not yet chosen a cobalt-steel magnet and a brass rod of equal weight. The magnet is a steel rod 5 inches long, $\frac{1}{8}$ in. in diameter and weighing 8 gm. Its determined magnetic moment is 570 c.g.s. units. The total weight is about 20 gm. and a phosphor-bronze strip suspension has been chosen of breaking strength 110 gm. wt. and torsional coefficient 0·024 dyne-cm. per radian. The calculated oscillation period of the system in the absence of magnetic effect is 330 seconds, which is of the same order as for gravity balances. Its sensitivity, which is determined by the ratio Ml/τ, where M, l, τ M is the magnetic moment, l the length of the beam arm, and τ the torsional coefficient of the suspension, may be anticipated to be equivalent to that of the early Eötvös balance. That is, with an optical magnification (easily attainable) of 2000, one scale division would correspond to a magnetic gradient of 10^{-8} gauss/cm., which perhaps we may call by analogy 10 Eötvös magnetic units*. With this instrument the horizontal variation of V at the magnetic equator would be just detectable. Greater sensitivity is clearly at our disposal, but I have thought it wise not to push this to the limit thus early.

And now for the practical difficulty. It arises from the fact that the magnetic axis of the magnet may not be quite vertical. If so, there will survive a small but troublesome horizontal magnetic moment, tending to twist the suspending fibre which in turn will transmit a spurious torsion to the central suspension. I have given a great deal of attention to the ways and means of elimination of this effect, so far unsuccessfully. Having regard to the data previously given we must rule out, I think, the possibility of adjusting the magnetic axis to be truly enough vertical. Nor can I see how to attach the fibre with sufficient precision to the magnet to allow it to rotate about its slightly tilted axis without raising or lowering its centre of gravity. (If this were possible, of course, all torsion in the fibre would automatically vanish.) Alternatively, is it possible to find some form of suspension, fibre or otherwise, in which there is support against gravity with negligible torsional control? Fibre

* The Eötvös (gravity) unit is defined as 10^{-9} dyne/gm.cm. or 10^{-9} sec.$^{-2}$

materials are roughly equal in respect of the ratio between torsional coefficient and breaking stress. No great advantage is gained by choosing one rather than another; and for none is the ratio small enough for the present purpose. It is a problem that must have cropped up already many times in the design of physical instruments*, and the absence of solution means, I fear, that the present difficulty may also remain unsolved. It is not so much a question of loss of sensitivity, as in the Cavendish form of the balance. That could be reduced well below fifty per cent, and I shall show later how, apart from the application to magnetic surveying in the field, the proposed balance promises remarkable accuracy in laboratory measurements of susceptibility. But magnetic surveying would require in normal practice certain changes of azimuth of the beam to be effected merely by the rotation of the instrument as a whole. Even if, as might easily be arranged, the torsion in the suspension were approximately zero in one azimuth, it would not remain so for other necessary azimuths of the instrument, and the indirect effect of H would be to produce large spurious deflections which could not be conveniently observed in the optical system designed to disclose the small effects of field distortion. Very precise manipulation of the torsion head would be necessary—as, indeed Eötvös himself found—constituting a serious complication in a field instrument†. The other advantages of the completely astatic system also would not be fully realized.

I have thus to leave at present undetermined the question of the practicability of this second form of magnetic torsion balance for field use. But before doing so I wish to give an illustration of its powers of detection, if only its construction could be achieved. Perhaps this will serve also as a justification for the persistency with which

* As, for example, in the fluxmeter. But we are dealing here with a question of precision of a very different order.

† One of the chief reasons for the accuracy of the Eötvös torsion balance for field observations is that the torsion head is *not* adjusted in the field. The optical system is capable of measuring only small angles accurately—up to a maximum of less than 1 degree. Using the notation given later on, page 11, we can see the way in which H gives rise to comparatively large deflections masking the field-distortion effect. Suppose we turn the instrument into a new azimuth. This means that α, the azimuth of the torsion head, is changed to α_0, and the geometrical relation

$$\alpha - \alpha_0 = (\theta - \theta_0) + (\phi - \phi_0) + (\epsilon - \epsilon_0)$$

holds. In a uniform field the equations of equilibrium are

$$\tau'\phi_0 = \mu H\epsilon_0, \qquad \tau'\phi = \mu H\epsilon,$$

$$\tau\theta_0 = \tau'\phi_0, \qquad \tau\theta = \tau'\phi,$$

whence $$\tau'(\phi - \phi_0) = \mu H(\epsilon - \epsilon_0) \text{ and } \tau(\theta - \theta_0) = \tau'(\phi - \phi_0).$$

Hence $$\alpha - \alpha_0 = (\theta - \theta_0)(1 + \tau/\tau' + \tau/\mu H)$$

or $$\theta - \theta_0 = \frac{\alpha - \alpha_0}{1 + \tau(1/\tau' + 1/\mu H)}.$$

This shows that $\theta - \theta_0$, the observed deflection, is not in general zero, as we should desire it to be under uniform field conditions. It is only if either τ' or μH is zero that $\theta = \theta_0$: in other words we must have one of three things—a negligibly small horizontal magnetic moment, a virtually zero horizontal field, or a practically torsionless subsidiary suspension. Without one of these conditions $\theta - \theta_0$ may be a large fraction of $\alpha - \alpha_0$, itself in practice amounting to π for one instrument-setting. For example, if μH is large and $\tau' = \tau$, then $\theta - \theta_0 = \frac{1}{2}(\alpha - \alpha_0)$, a large angle which cannot be measured in the field with sufficient accuracy to reveal precisely enough the small superimposed effects of slight non-uniformity of field.

mentally I have pursued the aim. Imagine a sphere of material of susceptibility K relative to its surroundings, buried at the magnetic equator with its centre at a depth below the surface equal to its diameter. Under the influence of the earth's field (here horizontal) it will become magnetized and produce at the earth's surface, vertically above, a small reduction ΔH of the earth's normal field, which proves on calculation to be given by

$$\Delta H = -0{\cdot}17K.$$

It is independent of the radius and depth of the sphere because the effect of depth is just compensated by change of magnetic moment with altered size. We may choose ΔH so as to represent the present practical limit of sensitivity of horizontal variometers in common use. This is about 5γ, and the corresponding value of K in our example is approximately 3×10^{-4} c.g.s. units. Turn now to the magnetic gradiometer described. The value of $\partial V/\partial s$ which it measures directly above the sphere is given
h by $3\Delta H/h$, where h is the depth of the centre; or with the same value of ΔH

$$\partial V/\partial x = 1{\cdot}5 \times 10^{-4}/h.$$

This equation makes evident an important distinction between the effects of buried magnetic deposits upon total change and upon gradient of magnetic intensity. The former does not vary with proportionate alterations of depth and linear dimensions of the deposit. On the other hand the gradient falls with increasing depth under the same conditions, as is shown by the appearance of h in the denominator. In the special case mentioned the magnetic force is inversely proportional to the cube of the depth; its gradient to the fourth power. This means that for the detection of large magnetic features at great depth the ordinary magnetic variometers, which measure ΔH or ΔV, will probably remain superior. At lesser depths, however, the advantage of the gradiometer becomes very marked. Thus, if our sphere is 15 metres in radius, and the depth of centre 30 metres,

$$\partial V/\partial s = 50 \times 10^{-9} \text{ gauss/cm.},$$

a value 50 times the anticipated sensitivity, under conditions in which ΔH would be only just measurable.

I pass now to consider the possible utility of the second magnetic torsion balance described, not as a field instrument but as a laboratory tool for accurate magnetic measurements. Here the main difficulty, I think, disappears. We accept the circumstance that the suspended magnet will in general have a small horizontal component of magnetic moment, causing it to tend to set with its axis in the plane of the meridian; also that in general this will give rise to torsion in the lower and central suspensions. This means that measurements must probably be made from a false zero. A diagram, figure 1, will illustrate the general angular relations; it relates to a plan of the apparatus. XX is the direction of H. AB represents the balance beam, the magnet being suspended at B and the counterpoise attached at A. The surviving horizontal component of the magnetic moment of the nearly vertical magnet is represented by cd, and ef indicates the azimuth of the torsion head above O. In the
θ, ϕ, ϵ ideal position in which all the angles α, θ, ϕ and ϵ are zero no couple acts on any part of the system, and both fibres are free from torsion. It follows that in the displaced

equilibrium position represented, θ is the torsion on the central suspension and ϕ that on the fibre immediately supporting the magnet. From geometry

$$\alpha = \theta + \phi + \epsilon.$$

If the system is first in a uniform field and then subjected to any additional clockwise couple G, without alteration of the torsion head, θ, ϕ and ϵ, but not α, will change so that we have

$$(\theta - \theta_0) + (\phi - \phi_0) + (\epsilon - \epsilon_0) = 0,$$

where θ_0, ϕ_0, ϵ_0 refer to the uniform field conditions (a) and θ, ϕ, ϵ to the modified conditions (b). θ_0, ϕ_0, ϵ_0

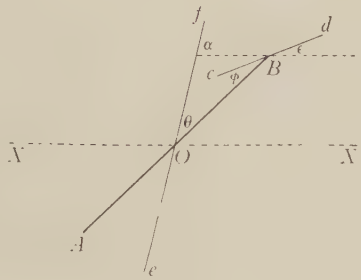


Figure 1

The equations of equilibrium for (a) are

$$\tau'\phi_0 = \mu H\epsilon_0^* \text{ for the suspended magnet,}$$
$$\tau\theta_0 = \tau'\phi_0 \text{ for the beam,}$$

where τ and τ' are the torsional coefficients of the central suspension and subsidiary fibre respectively, and μ is the horizontal magnetic moment, small in our case, but not necessarily so in the argument. τ, τ' μ

Under conditions (b) we have

$$\tau'\phi = \mu H\epsilon \text{ for the magnet,}$$
$$\tau\theta = \tau'\phi + G \text{ for the beam.}$$

On subtracting the corresponding equations we obtain

$$G = \tau(\theta - \theta_0) - \tau'(\phi - \phi_0)$$

and

$$\tau'(\phi - \phi_0) = \mu H(\epsilon - \epsilon_0).$$

Using the geometrical relation of angles first given, $\phi - \phi_0$ and $\epsilon - \epsilon_0$ may be eliminated, leaving G in terms of $\theta - \theta_0$, thus,

$$G = (\theta - \theta_0)\left\{\tau + \frac{\tau'}{1 + \tau'/\mu H}\right\}.$$

The quantity which can be observed optically is $\theta - \theta_0$, although θ_0 itself is unknown. It is the change of deflection due to applying the additional couple G. The sensitivity of the arrangement is measured by $\theta - \theta_0/G$, and the effective torsional coefficient, which we desire to be small, is the rather awkward expression

$$\tau + \frac{1}{1/\tau' + 1/\mu H}.$$

* It appears in practice that ϵ will always be small, so that $\sin\epsilon = \epsilon$ nearly enough.

The positive addition to τ shows the diminished sensitivity arising from the existence of a horizontal field on the one hand, and the incomplete elimination of horizontal magnetic moment on the other. If μH were negligible the second term would vanish, and we should secure the maximum sensitivity represented by the reciprocal of τ. But even at the other extreme things are not so bad; for as H increases indefinitely the second term approaches τ' and the maximum effective torsional coefficient is $\tau - \tau'$. The magnet suspension can easily be assigned in practice a smaller torsional coefficient than the central suspension, for it has to support a smaller weight. Accordingly the sensitivity attainable is more than half the ideal.

A concrete illustration may be given to show how this still high sensitivity promises to facilitate accurate measurement. Consider the magnetic gradiometer already described, and let us assign to the fibre supporting the nearly vertical magnet a torsional coefficient equal to that of the main strip, namely, 0·024 dyne-cm. per radian. The strength of each of the poles of the magnet is about 50 c.g.s. units. Imagine a large (effectively infinite) slab of material placed with its vertical face
d parallel to the plane of the system so that the poles are distant d from the surface. The attraction on each pole arising from magnetic induction will be

$$f = \pi K m^2 / 2d^2,$$

K, m where K is the susceptibility of the material of the slab, and m the pole-strength. With d equal to 1 cm.—a geometrically practicable condition—the total horizontal
f translational force on the magnet is $2f$ where

$$2f = \pi K m^2 / d^2 = 8 \times 10^3 K.$$

G The couple G acting, the beam arm being 2 cm., is given by

$$G = 1{\cdot}6 \times 10^4 K$$

and the observed deflection $\theta - \theta_0$ by

$$\theta - \theta_0 = G/(\tau + \tau') = 1{\cdot}6 \times 10^4 K / 0{\cdot}048 = 3{\cdot}3 \times 10^6 K.$$

With K assigned even so low a value as that of air, namely $3{\cdot}2 \times 10^{-8}$,

$$\theta - \theta_0 = 0{\cdot}01 \text{ radian},$$

an angle easily made evident by ordinary optical means as a deflection of several centimetres. To realize approximately the proposed infinite slab of material we should require a suitable plane-faced vessel which could be evacuated at will. The entry and exit of the air or other gas could then be observed magnetically, and the possibility of somewhat precise measurement of susceptibility is apparent, under conditions in which the magnetizing fields are not large. The method would be a kind of inverted Curie balance, the magnet, instead of the test material, riding on the movable beam.

Finally I wish to add a few remarks on the question of the possible elimination of the effects of the earth's horizontal field. I do not think this would be practicable in a portable field instrument, on account of the size of the accessory apparatus which would apparently be necessary. But there is reason to believe that something could

be done in the laboratory to avoid the appreciable loss of sensitivity before mentioned, and to escape working from a false zero with its implications of temperature variability. Why should we allow the earth's field to exercise so much constraint? In my thoughts on this problem I have dreamt of a small laboratory which is devoid of magnetic force except from sources deliberately introduced. But, leaving that aside as too ambitious, we may consider how far a comparatively small space—large enough for a torsion balance of the size indicated—may be rendered magnetically free, at any rate in respect of horizontal force. With H small, as well as μ the horizontal magnetic moment, their product may be negligible in comparison with τ', the torsional coefficient of the magnet suspension. This would imply that the sole control is exercised by the central suspension, which would attain a true zero of no torsion in any azimuth of the instrument as a whole, except when disturbed by the applied forces which we aim at measuring. The problem really amounts to this—Can we by artificial means cancel the earth's horizontal field to a sufficient extent?

μ
τ'

I have made some calculations in this connection, based on available data for the materials and properties of magnets and suspensions. The assumption is also made that it would be possible to suspend the magnet with its axis so nearly vertical that the horizontal component of moment does not exceed 1/500 of the total. This could be secured, I think, without too much difficulty. On this basis, if we could reduce the horizontal field to one-thousandth part of its natural value in London the surviving magnetic control would not be more than 1 per cent of that provided by the torsion of the central suspension. With more complete cancellation of the field its influence is, of course, correspondingly diminished. Under the conditions specified, also, there is no notable advantage in making the torsional coefficient τ' small, except in so far as suspension of the magnet by a flexible fibre rather than a more rigid attachment would probably facilitate the necessary adjustment of the magnetic axis to the vertical.

With the cancellation of H in view I have been making preliminary experiments with a pair of coaxal current-bearing coils adjusted to the proper azimuth. I believe that it will be possible to maintain over the current a control sufficiently precise to reduce the effective horizontal field for considerable periods to one-thousandth of the natural field, or even less. So far progress has been slow because of the variability of the magnetic field in the precincts of the Imperial College while London is electrically active. And even in the quiescent period, from 1 a.m. to 4 a.m. approximately, one is liable to be disturbed, as I have found, by magnetic storms. However, it is merely a matter of choice of suitable place and time to determine the practical possibilities of the arrangement, which will be described in due course.

I may perhaps refer to one aspect of the matter which is of some interest. The coils were designed to satisfy a condition of field-uniformity of 1 in 10^4 at 10 cm. from the centre of the pair arranged in the Helmholtz manner. This implied rather large dimensions—a diameter of about 2 metres. From motives of economy the formers were made of wood, and as it was considered easier in our workshop to make hexagons than circles of this size, the former shape was chosen. I am indebted to our carpenter, Mr Leech, for making these hexagonal formers accurate in all

respects to approximately two parts in a thousand, without the use of special tools. The choice of this shape necessitated the examination of the condition of separation appropriate to maximum field-uniformity near the centre of the pair. This led to a more general investigation in relation to plane regular polygons of any number of sides arranged coaxally with corresponding corners opposite one another. Several unexpected results appeared in the analysis. The first is the small variation of the distance of separation of the coils for maximum field-uniformity resulting from change of coil-shape. If we denote by b half this optimum distance, and by a the perpendicular from the centre of the polygon to any side, the ratio $a\ b$ has the value $\sqrt{3}$ for one extreme, namely, the polygon of 2 sides, and the value 2 at the other limit, the polygon with infinitely numerous sides. The first case, of course, corresponds to two pairs of very long wires, all parallel, and forming a rectangle in cross-section; while the latter is clearly realized in circular coils. For all other regular polygons a/b lies between $\sqrt{3}$ and 2, or $a^2\ b^2$ lies between 3 and 4. I append a table, which may possibly be of use in other connections, giving a number of intermediate values.

b, a

Table 1

Number of sides of polygon	$\frac{a^2}{b^2}$	$\frac{a}{b}$
2	3·00000	1·73205
3	3·12153	1·77086
4	3·37284	1·83653
5	3·55561	1·88563
6	3·67454	1·91691
8	3·80725	1·95122
∞	4·00000	2·00000

With regard to the degree of field-uniformity at the appropriate coil-separation, the circular shape is best for a constant length of wire, as regards both the variation along the axis and that perpendicular to it. But the superiority in this respect is not great. The hexagon, for example, although inferior to the circle escribing it, is better than the inscribed circle. For the same perimeter the axial-field variation near the mid-point is only about 25 per cent greater for hexagons than for circles, and amounts to just less than 10^{-4} of the mid-point value at 10 cm. from the mid-point of the coils constructed, in which the perpendicular distance between opposite sides is 2 metres. A somewhat surprising result is with reference to the field variation in the medial plane perpendicular to the axis. It proves to be the same in the directions of the diagonals of the hexagons and of the perpendiculars between opposite sides. Moreover, its amount is exactly $\frac{3}{8}$ of the axial variation, just as Nagaoka* has shown it to be for circular coils arranged correspondingly. I hope to give a more complete description of this interesting field on a more appropriate occasion.

I must ask your indulgence for presenting to you a subject which remains so largely unfinished from the experimental point of view. On the theoretical side it has involved much work, and as I said at the beginning I have deliberately adopted the brain-saving device of addressing you upon the things to which spontaneously my thoughts have turned so often since I became your President.

* See F. E. Smith, *Phil. Trans.* A, **223**, 187 (1923).

Note added November 22, 1933

My statement on p. 3 that the Eötvös torsion balance has had no opportunity of displaying its virtue of immunity from the effects of temporal variations of gravity is incorrect. When I made it I had forgotten that there are in fact such fluctuations, the chief of which is a semi-diurnal one due to the moon. These are too small to be revealed by ordinary pendulum observations, but Loomis* measured the lunar effect for the first time in 1931, by comparing the rates of Shortt clocks and quartz oscillators. The complete absence of response of the Eötvös instruments in the extensive field practice of the last forty years proves in a most convincing manner that they exclude time variations of gravity and reveal space variations only. I propose later to communicate a short note giving some details of quantitative comparison between the small effective space changes and the large but ineffective time changes of the terrestrial gravitational field.

* *M.N.R.A.S.* **96**, 569 (1931).

75.023.2:612.843.34

THOMAS YOUNG AND THE SIMPLIFICATION OF THE ARTIST'S PALETTE

The Thomas Young Oration delivered on October 6, 1933 by

HERBERT E. IVES

§ 1. INTRODUCTION

In the year 1807 Thomas Young published in his *Lectures on Natural Philosophy* his observation that all the colours of the spectrum, and consequently all colours occurring in nature (since these are mixtures of spectrum colours) may be reproduced by mixtures of three—red, green and blue, or blue-violet as it is sometimes termed. Young represented his findings in the form of a colour triangle which is shown, carefully coloured by hand, in plate XXIX of Volume 1 of his *Lectures*. Much work in the way of scientific amplification and elucidation of Young's idea has been done since his time, notably by Maxwell and by Helmholtz, and the facts of three-colour mixture are the common property of scientific students of colour. In the field of technical application, three-colour photography and typographic printing have made extensive use of Young's discovery.

At the time Thomas Young published his three-colour theory, the great painters of the late eighteenth and early nineteenth centuries, Gainsborough, Reynolds, Romney, Stuart, Raeburn, and others, were painting their pictures with a set of pigments or "palette" which they had inherited from the old masters. The only theory involved in this palette was the primitive one of having a permanent and reliable colouring material approximating to each colour that the artist might want to reproduce. This primitive idea still dominates the artist's palette, for while the mixing of pigments is a common and unavoidable practice in the painting of pictures, the number of pigments in use is still very great. A. P. Laurie in his *Painters' Methods and Materials* lists some thirty odd reliable pigments for oil painting, and adds, "There are other pigments to be found in artists' colourmen's lists. . . but their name and number is legion." The average artist uses from twelve to twenty—Ingres used twenty-seven. Some artists, who have specialized in one line of work, are held up to admiration for the small number of colours with which they manage to get along—for instance Gilbert Stuart, who did his portraits with only seven colours. Usually, however, the work of artists who employ a limited palette is characterized by a preference for certain colours and colour combinations to which they are driven by the limitations of their chosen few paints. With such a vast array of pigments, greatly varying in their spectral absorption characteristics, it is a necessary consequence that there are many possible ways of achieving certain colours by mixture,

and also that some pigments will not mix to yield useful colours*. There can of course be no simple guiding theory covering the mixture of such a large array of pigments of all hues and varying degrees of saturation. Learning how to use his colours becomes a matter of rule-of-thumb experimentation and practice for each individual painter.

With the general scientific recognition of the great simplicity introduced into problems of colour-mixture by the observation of Thomas Young, it is a matter for surprise that no reflection of his work has appeared in the craft of the painter. There appear to be several reasons for this; the dominating one is probably that no professional scientist familiar with three-colour theory has been interested in painting. Conversely no competent artist has been sufficiently at home in optics to make a practical application of the three-colour principle. There exists among artists and teachers of art a belief, which is occasionally expressed very emphatically, that there is no connexion between the phenomena of light mixture and the practice of pigment mixture. Such attempts to use three-colour theory as are on record have failed for two reasons:

(1) Where it has been tried by artists there has not been a correct understanding of the transition from the *additive* mixture of red, green and blue light, to the *subtractive* or absorptive mixture which demands colours complementary to these, namely, a minus red (blue-green), a minus green (crimson or purple), and a minus blue (yellow). A long-perpetuated error has designated as pigment primaries red, yellow and blue. These, if pure, are three colours which could not be mixed to give by absorption the whole range of spectral hues. A number of artists have attempted three-colour painting, using approximations to pure red, yellow and blue, such as vermilion, chrome yellow and ultramarine, but with results which have been interpreted as discrediting the three-colour theory for use in painting. An explanation commonly given is that the pigments available are not sufficiently pure in colour. This, however, is an entirely erroneous idea, for it is only by their "impurity" that pigments mix to give other colours. Yellow and blue will give green only if they have overlapping regions of reflecting power, that is, if they are both impure. Reasonably pure red and blue pigments, having no overlapping spectral reflection regions, would simply give black, although the statement, carried over from additive colour-mixture, that red and blue make purple, is often found in artists' texts.

(2) As we shall see, even had anyone interested in the use of three pigments planned to use the proper colours, he would have been seriously handicapped until recently by the fact that reasonably correct colours have not long been available in satisfactory pigment form. He could not have attained success without the co-operation of the chemist in developing pigments which were unknown to the old masters or the contemporaries of Young.

* Thus the article of faith of many painters that "yellow and blue make green" is upheld by a mixture of lemon-yellow and Prussian blue, which have wide reflecting regions in the spectrum, leaving a band of residual reflection of green, but is contradicted by a mixture of ultramarine and chrome yellow, whose reflection bands scarcely overlap, so that the resultant green is quite dull.

§ 2. THREE-COLOUR PAINTING BY THE ADDITIVE METHOD

The most illuminating approach to the understanding of the principles of three-colour painting may be made by taking up first a theoretically interesting but impracticable method, which exactly parallels the mixture of colours by the additive mixture of three coloured lights. Let us reconsider, in somewhat more detail, the colour triangle of Young. Young drew this triangle as a straight-sided triangle lying in a plane, with red, green and blue at the vertices, and white (the mixture of red, green and blue) in the centre. Along the red-to-green side of the triangle lie orange, yellow and yellow-green, along the green-blue side the blue-greens, and along the red-blue side the various crimsons and purples. Mixtures of white with the colours on the edges of the triangle, or tints, lie at intermediate positions between the white centre and these colours. Later researches have shown that the true representation of the spectrum is not a triangle with straight sides, but one with slight curvature and a very blunt apex at the green corner. In terms of colour-mixture, this means that the mixtures of spectrum red, green and blue give colours which, while matching the rays of the spectrum in hue, are somewhat less saturated. The complete diagram is also not a triangle but a solid figure, of which Young's triangle is a cross-section.

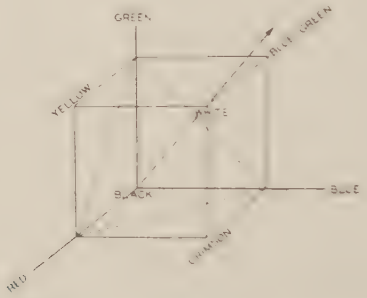


Figure 1. Trichromatic mixture phenomena represented by a colour cube.

This complete three-dimensional presentation of the phenomena of three-colour mixture is shown in perspective form in figure 1. Our figure is bounded by three straight-line Cartesian axes. The origin is the condition of no light, or black, the three axes represent red, green and blue light. Thus any line parallel to the red axis indicates addition of red light, any line parallel to the green axis addition of green light, and any line parallel to the blue axis addition of blue light. By proceeding along any one axis a chosen distance, then at right angles thereto, and then at right angles again for the same distance, one arrives at a point on the diagonal axis of the solid angle where the three colours are present in equal quantities, the black-white or grey line. The several possible paths made in this way define a cube, three of whose corners are red, green and blue; another three corners are the mixtures of red, green and blue two by two, and the remaining two corners are black and white. All possible colours, including tints and shades, are represented by points within this cube. The colour triangle of Young is the section of the cube, shown dotted, through the red, green and blue corners. Its centre is grey, and the spectrum lies

along its sides. If we imagine ourselves looking at this triangular section along the black-white axis of the cube, and imagine the cube as a hollow skeleton, we obtain the view shown in figure 2 *a* with the black origin at the centre. The primaries, indicated by arrows, to be used to make mixtures matching any colour can be picked by inspection, as those between which the colour in question lies. Interpreted in terms of the mixtures of three lights, red, green and blue, this cube tells us that we can project upon an unilluminated white (or black) screen, red, green and blue in various amounts and thus make all colours, with the reservations as to saturation already noted.

Let us now see how to apply these experimental facts to putting down our colours, not with light but with pigments. For this purpose we must start with the equivalent of the unilluminated screen, or in other words with a black surface. Our colours as we put them down must be red, green and blue, and should be opaque or capable in sufficient thickness of completely reflecting all the incident light lying within their range of spectral reflecting power. We cannot superpose these colours as we can superpose lights, for if transparent they will be mutually destructive, and if they were opaque only the top one would count. We must, therefore, place these down in

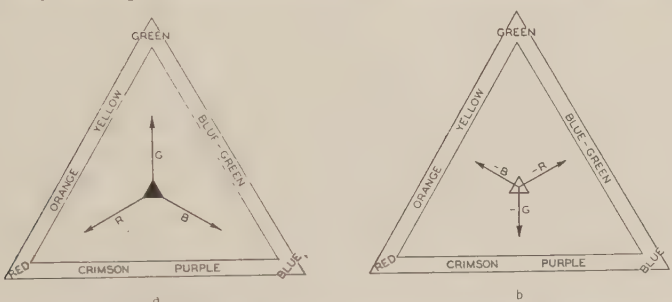


Figure 2. The section of the colour cube known as the colour triangle: (*a*) With the rear vertex of the cube exposed, to illustrate additive colour mixture; (*b*) With the front vertex of the cube displayed, to illustrate subtractive colour mixture.

juxtaposed small spots to be mixed by retinal fusion, and they should be chosen, for convenience, of such reflecting power that a group of spots of approximately equal area will together give the impression of grey or white. In order for the mixed colours to be as pure as possible the red, green and blue pigments should be of high purity of colour, that is they should reflect comparatively narrow spectral regions.

The technique of painting with these pigments consists in starting with a glossy black ground, such as photographers' ferro-type plate, and spotting on the three pigments in as small areas as is practicable, using large completely juxtaposed spots for the whites, and reducing the size of the spots, that is, leaving larger black areas between them and varying their relative proportions, to give all the mixed colours. Black, of course, remains where the ferro-type plate is untouched.

It is obvious that the technique of painting pictures by this method is very slow and difficult, and from that standpoint alone would have nothing to recommend it. There is, however, another unescapable defect of this type of picture, which is its inherent low luminosity or reflecting power. The brightest white which can be

obtained is that given by an area completely covered with juxtaposed pigment spots, whose individual reflecting power is extremely low and is lower the purer the pigments used. The reflecting power of such a white made with available pigments approximating to a monochromatic character will hardly exceed 10 per cent. A picture of this sort hung alongside ordinary pictures in which the whites have a reflecting power of 80 per cent or thereabouts, or in a room in which the whole range of brightness of coloured objects lies between the blacks and whites of 80 per cent reflecting power, appears hopelessly dull. If, however, one arranges to illuminate one of these three-colour pictures to approximately ten times the illumination of any of its surroundings, this handicap can be overcome*. For this purpose the picture should have a delimited spot of light such as can be produced by projecting the field of a lantern-slide projector upon it and surrounding the picture with an absolutely black area, for instance by mounting it in the centre of the open side of a box lined with black velvet. In order to achieve success in this scheme of illumination, the black ground must be of negligible reflecting power, which is approximately achieved by using a specular black surface such as the ferro-type plate, and the pigment spots should be laid on very flat so as to give no bright spots of specular reflection. The projection lantern should also, in order to avoid any specular reflection from the ferro-type plate, direct its light at a considerable angle to the surface of the picture from below or above. By utilizing this very special scheme of illumination one can carry through this method of painting in three colours, which closely exemplifies the additive three-colour mixture phenomena with coloured lights. It is obviously only of theoretical interest†. In passing we may recall that this kind of additive three-colour picture is illustrated by the autochrome process, in which the loss of light is one of the major defects.

§ 3. THREE-COLOUR PAINTING BY THE SUBTRACTIVE METHOD

The methods of painting which have come into practical use depend not upon the addition of colours to black, but upon the subtraction or absorption of colour from white. The final colours are due to residual reflecting power after the pigments are mixed together. An artist starts ordinarily from a white canvas and not from a black ferro-type plate. He starts also from a white which is of the same high reflecting power as the whites in the objects which will be in the same possible fields of view as his picture. It shall be our problem now to investigate how the three-colour principle may be applied to painting upon a white canvas by the process of absorbing light.

As a simple approach to this problem from the ground-work already set up, let us consider subtractive three-colour painting, starting not with the usual white canvas of high reflecting-power reflecting the whole spectrum, but with a white of

* This intensified illumination of course increases the surface reflections in the same ratio, and makes it difficult in practice to prevent a general greying over of the picture.

† The method used in this kind of painting was to a certain extent utilized by the "pointilist" school, who put down their colours in juxtaposed spots, and depended on retinal fusion or "vibration of colour" for mixture. The painters of this school, however, did not select red, green and blue as their colours, and used very unsaturated colours to get their greys and whites of sufficient luminosity.

the type considered in the last section, namely a surface uniformly covered with small juxtaposed areas of red, green and blue. This will appear grey alongside an ordinary white canvas, but, as has already been described, might for experimental purposes be raised to white by localized illumination. Let us now imagine that we provide ourselves with a pen charged with opaque black ink, such as India ink, and that we go over the red, green and blue spots and reduce them individually to spots of smaller and relatively different areas in different parts of the picture. Where we completely block them out we get black, where we block out the blue spots we get red plus green, that is yellow, where we block out the green spots we get purples, and where we block out the red spots we get blue-greens. If we are sufficiently patient and skilled, we can again produce a picture exhibiting all colours; here again, by a very impracticable process, which might be described as three-colour mezzotint. Referring to the colour cube of figure 1 our process consists in working back from the white vertex along directions parallel to the three axes, going in the reverse directions from those considered in the case of additive painting.

Let us consider now a variation of the procedure just described. Instead of using a finely pointed pen as our implement and confining ourselves to black ink, let us go one step nearer the technique of the artist and use brushes. Let us use three brushes each charged with a transparent absorbing paint such as water colour. One of these paints should have an absorption in the spectrum such that it obstructs or absorbs only the light reflected from the red spots, the second should absorb only the light from the green spots, and the third should absorb only the light from the blue spots. If now we take our three brushes and lay down on various parts of the stippled white surfaces various proportions of these three washes, we can with free brush stroke (thus getting away from the hampering microscopic spotting of the surface heretofore employed) proceed to paint a picture.

But an even simpler procedure is open to us. Instead of putting down the colours in three washes, let us mix them beforehand on a palette which is itself a red, green and blue stippled surface. We shall thus see upon the palette, as we mix our three washes, the colour which they will produce when laid upon the canvas. In this way we can mix and lay on (transparent) colours in practically the same manner as does the water-colour artist who works in aquarelle. Our resultant picture is subject to the same limitations as to brightness and necessary conditions of illumination as the additive one we have previously produced.

Before we go on to the problem of painting on a truly white canvas, there is one more approximation we could make to the technique of the painter in oil colours. A great working advantage of oil paints is that one is emancipated from dependence upon the white canvas for whites or admixtures of whites. The artist can at any time restore the canvas, as it were, by mixing white pigment with his colours and painting over earlier work at any stage. We could parallel the regular oil-painting technique with the subtractive process as thus far described by providing ourselves with a white pigment which reflected only three narrow spectral regions just as does the special canvas we have used. This point is brought up merely to aid in introducing the complete technique of three-colour painting in oil to be discussed in detail.

Let us consider the appearance and characteristics of the three colours and the white which we have just imagined ourselves using to build up a picture by the subtractive or absorptive process. The white is, of course, what would ordinarily be described as a dark grey. As we apply each one of the colours alone over the grey, we note that their colours are complementary in hue to the additive primaries red, green and blue. The colour which absorbs the red leaves a mixture of blue and green, that is a blue-green. The colour which absorbs the green leaves a mixture of red and blue, which is a crimson or purple; the colour which absorbs the blue is a mixture of red and green, or yellow. When we use these colours in pairs we obtain the primaries for mixtures of light, that is red, green and blue. When we use all three, we obtain black.

By way of review and crystallization of the ideas which have just been developed let us refer to figure 3. In diagram *a* we have represented the additive mixture of red, green and blue, building up to a trichromatic white, of low reflecting power. In diagram *b* we have represented the subtractive method as just described, starting with a trichromatic white and arriving at black by absorption.

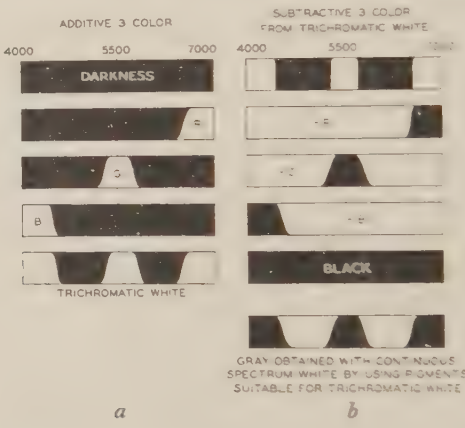


Figure 3. Additive and subtractive three-colour mixture illustrated.

Now the colours which we have used in this subtractive method are unspecified as to the extent of their spectral absorptions. They must merely absorb one of the additive primaries and nothing of the other two, but as to their behaviour in the spaces in the spectrum between the tri-colour white it has not been necessary to make any restrictions, for these spaces are non-reflecting and contribute nothing to the final result. Let us now go over to the practical case of the use of a truly white canvas, with an ordinary white opaque pigment such as zinc oxide for restoring the canvas or for admixture with the coloured pigments to reduce their saturation. These whites, unlike those trichromatic whites considered above, reflect the whole of the visible spectrum. A question fundamental to our problem is this: starting with a continuous-spectrum white, can we postulate three absorbing media which, as their mixing proportions are varied, will yield the same range of colours as absorptions applied over the trichromatic white? Specifically can we imagine three pigments

which (1) together absorb the whole spectrum to give black; (2) in pairs give satisfactorily pure red, green and blue; and (3) singly give yellow, crimson and blue-green matching the mixtures of pure red, green and blue in pairs?

Taking up these specific requirements in turn we note first, by reference to the lowest sketch of figure 3 *b*, that the localized absorptions which sufficed with the trichromatic white will not together cover the spectrum and give black. *For this purpose the absorption bands of the three pigments must meet or overlap.* If they merely meet, it will be possible to produce a black but the three primaries obtained by using them in pairs will consist of broad sections of the spectrum, instead of the practically monochromatic red, green and blue we want. In order to secure primaries of high purity (our second requirement) it is therefore necessary to have absorptions which overlap. At the same time the non-overlapping portions of any two must be free from absorption so that the residual colours shall be as bright as possible. Such absorptions are shown in the colour triangle in figure 4. As to the third requirement

Figure 4. The absorption regions of the three subtractive primaries represented in the colour triangle.

above listed we find the desired property to be guaranteed by the colour triangle construction. Upon plotting our three absorptions meeting the first two requirements upon the colour triangle, as shown in figure 4, we see at once that the minus red, minus green and minus blue constitute the flat sides of the triangle, that is, the extended portions of the spectrum which match in hue the mixtures of their two ends. In other words the additional light admitted from the continuous-spectrum white as the three absorptions are varied in extent or intensity is of a hue which is always of the kind we desire to produce by mixtures in the case of the trichromatic white.

We have thus far considered merely the question of hue, finding that we can handle it satisfactorily with a set of minus colours. In addition there are the factors of luminosity and saturation to be studied. In the cases we have considered where trichromatic whites were used, luminosity and saturation were automatically taken care of because the same elements entered into the whites and the colours. Can we, with the absorptions rendered necessary by the basic continuous-spectrum white, make mixtures which preserve luminous values, as do additive or subtractive three-colour mixtures using a trichromatic white? The answer is in general in the negative. The fact we must recognize is that whenever, by the subtractive method here considered, a subjective colour match is made, different in spectral composition from the sample, different portions of the spectral luminosity curve are affected so that a simultaneous hue and luminosity match is prevented.

This behaviour of subtractive mixtures of the type we are considering is made evident by applying the information contained in the three trichromatic excitation

curves and the luminosity curve of the spectrum, figure 5. As a typical illustration consider a lemon yellow, owing its colour to a narrow but intense absorption in the extreme blue of the spectrum, as shown in the dotted areas of the excitation (A) and luminosity (B) curves of figure 5 and in the spectrophotometric plot (C). Let us now endeavour to match this by using, with white, absorption bands of the kinds shown in figure 4. Let us assume that, upon admixture with white, these absorption bands remain of constant width, merely changing in intensity. Our process of making a match in hue would then consist in first mixing a certain amount of minus blue (yellow) with the white as shown by the cross-hatched area to the left in figure 5 C. This absorption would extend well into the green excitation curve of figure 5 A,

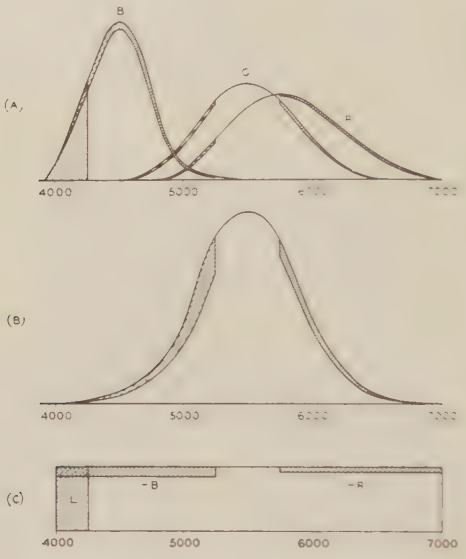


Figure 5. Luminosity and saturation problems in subtractive three-colour mixture illustrated.

making the resultant colour too orange. To offset this the red excitation curve will have to be reduced by the admixture of some minus red, as shown by the cross-hatched areas to the right. By this means the balance of red, green and blue excitation curves is restored, that is, a hue match is obtained. However, on reference to figure 5 B, where the same absorptions are represented in luminous values, it is at once evident that we have obtained our hue-match by encroaching upon the central more luminous portions of the spectrum; our mixed colour is too dark.

From this illustration, which is more or less representative, we see that subtractive mixtures obtained in this way from a continuous spectrum white, running darker than natural colours obtained from narrow intense absorptions, will drive us to paint in a lower key than a palette consisting of an extended range of pigments might permit; or, as an alternative, to raise the luminosity by greater admixture of white, thereby reducing the saturation. If the latter is done, then to the loss of

saturation already noted as inherent in trichromatic mixture is to be added a loss of saturation due to the use of a continuous-spectrum white*.

The situation is not, however, as serious as it might appear from the above illustration. Actual pigment absorption bands do not decrease merely in intensity on dilution with white. They simultaneously become narrower, in the manner shown in figure 6 where three idealized pigments are postulated, conforming in their concentrated form to the requirements given in figure 4, and behaving on dilution in a manner which is typical of pigments which might reasonably be expected to be found or produced. With such absorptions matches with colours due to narrow absorption bands will be made to a considerable degree by an actual narrowing of absorption bands in the mixtures, and so the need for lowering the key will largely be eliminated. In addition to this alleviating fact is the very general phenomenon

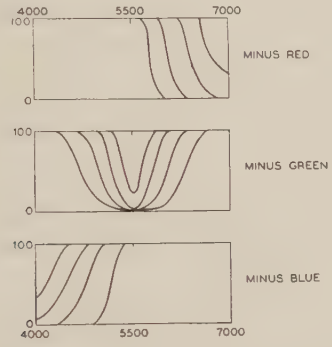


Figure 6. Ideal reflecting characteristics for three-colour pigments as diluted with white.

that all colours encountered in naturalistic painting are themselves quite unsaturated, in other words, the kinds of colours which give trouble in three-colour subtractive mixture are of rare occurrence. Also, if we are comparing a three-colour palette with the existing painter's palette, we must not forget that very few of the reliable pigments now used exhibit narrow, sharply defined, or intense absorption bands. In fact, the limitations which have been conscientiously pointed out in the three-colour palette are present also in large degree in the ordinary artist's palette. We have therefore a reasonable expectation that a simple three-colour palette of absorbing media of the kind specified by figures 4 and 6 will make satisfactory mixed colours for all ordinary painting. If, on the other hand, it proves deficient in practice, we are in an excellent position to know how to modify it in a scientific manner to meet all requirements.

We are now ready to consider the experimental test of the theory. This test consists in obtaining or producing three pigments approximating in satisfactory degree to the characteristics depicted in figure 6, and trying practical painting

* With such absorbing media as are actually available other faults, such as spreading of the absorption over spectral regions which should be clear, incomplete absorption at the centre of the absorption band, lack of proper transparency, etc. are probably much more serious.

experiments with them. Mixed together in varying proportions the three pigments should satisfactorily reproduce the hues of the spectrum. From these spectral matches mixtures with white will then give all the tints. Further mixture with black (combination of all three pigments) will then give all shades.

§ 4. PRACTICAL TEST

Although the general scheme for the simplification of the pigments used in painting, which has just been described, is applicable to any form, such as water colour or fresco, I have confined my attention to oil painting as the most powerful and flexible method. This study naturally divides itself into two stages; first is the stage of finding or producing a set of three pigments which shall have the proper character of absorption, in order to prove the correctness and advantages of the idea. The second stage, granted the success of the first, is to obtain or produce pigments of the required character which shall have other characteristics, such as fastness to light and freedom from chemical interaction, which are indispensable if the pictures painted are to have satisfactory permanence.

In regard to colours which shall be capable of proving the theory, certain characteristics other than those specified by the character of the absorption are desirable. It is, for instance, quite necessary, if pure brilliant colours are to be produced in the final picture, that the pigments give as little admixed white light by surface reflection as possible. This means that the pigments should have very intense colouring power and should have a considerable degree of transparency. They must, however, because of the nature of oil painting, be sufficiently opaque in concentrated form to ensure that under-painting shall be completely obscured.

On a preliminary survey, the colours called for appeared to be quite far from any available among the common artists' pigments. Accordingly an extended search was made, largely by spectro-photometric analysis*, in the field of organic dyestuffs. As a result of this search I have been able to select a set of three colours of a high degree of permanence which approximate fairly satisfactorily the ideals above postulated. These are, for the blue-green a tungstic acid lake of the basic dye Glacier blue; for the crimson a tungstic acid lake of Rhodamine 6-G; for the yellow the pigment known as Hansa yellow 5-G. The reflection spectra of these, with various dilutions with white, are shown in figures 7, 8 and 9. It will be seen that the yellow is a very satisfactory approximation to the ideal postulated; the crimson is fairly good, its defect being too low a reflecting power at the ends of the spectrum, particularly in the blue. The weakest member of this series is the blue-green; it is of the proper hue, but reflects some deep red, is of comparatively low reflecting power in the blue and violet of the spectrum, and exhibits, in general, too narrow a band of reflection; it approximates to a spectral blue-green rather than a minus

* I am indebted to Prof. A. C. Hardy of the Massachusetts Institute of Technology for making the very large number of measurements by means of his photoelectric recording spectro-photometer. The curves shown are all in terms of magnesium oxide as a standard white.

red*. Nevertheless, in spite of their obvious deviation from the ideal, this set of colours proves to be remarkably successful in copying most of the commercial artists' pigments and is notably better than any of these latter for producing certain colours such as purples. Because of the low reflecting power of the blue-green lake

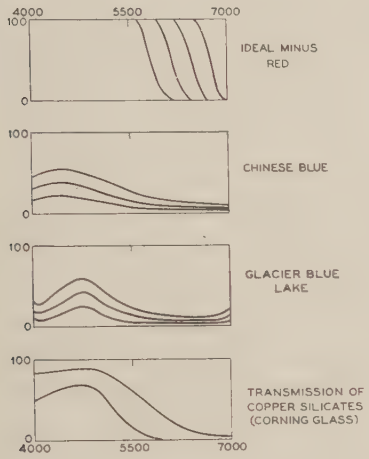


Figure 7. Ideal and available minus reds.

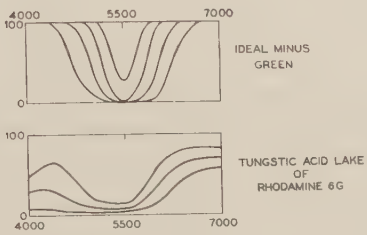


Figure 8. Ideal and available minus greens.

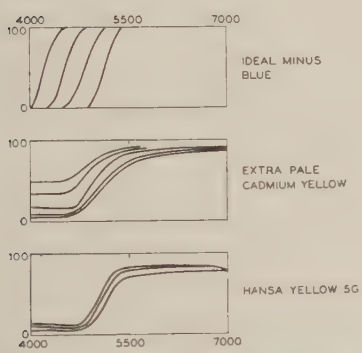


Figure 9. Ideal and available minus blues.

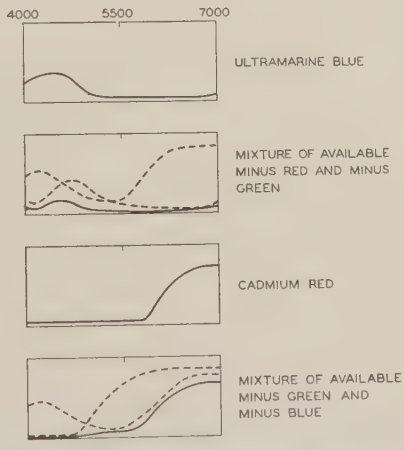


Figure 10. Illustrative mixtures obtained with the experimental three-colour palette.

* It is of interest to note that the colouring-materials commonly used for yellow are all true minus blues—the whole spectrum diminished by absorption in the blue—and are hence of high luminous value. On the other hand, all pigment blue-greens, with the single exception of cerulean blue (cobalt-tin-magnesium oxide) approximate toward monochromatic blue-green, and are hence of low reflecting power as well as being unsuitable for a subtractive three-colour pigment. A recurring difficulty in attempting to secure minus-red pigments in these experiments has been that all practical pigment makers instinctively turn to some mixture of blue and yellow-green, such as cobalt blue and chromium oxide, and have trouble in grasping the idea that the same hue should be obtainable by mere absorption in the red. The parallel situation in the case of yellows would be, if these were habitually produced by mixtures of orange and Prussian blue, best described as dirty browns.

in the blue, brilliant blues such as are given by ultramarine are not well rendered. This, however, is nothing to the discredit of the underlying theory as the reason is perfectly evident from the spectro-photometric curves. Figure 10, which gives spectro-photometric curves of representative mixtures, illustrates the unsatisfactory character of the blues, as just noted, and the excellent spectroscopic match with a common artist's pigment red*. With this set of colours it is possible to paint a representation of the spectrum of great fidelity, affording a convincing testimony to the validity of the theory. This set of pigments suffices to prove the soundness of the three-colour method.

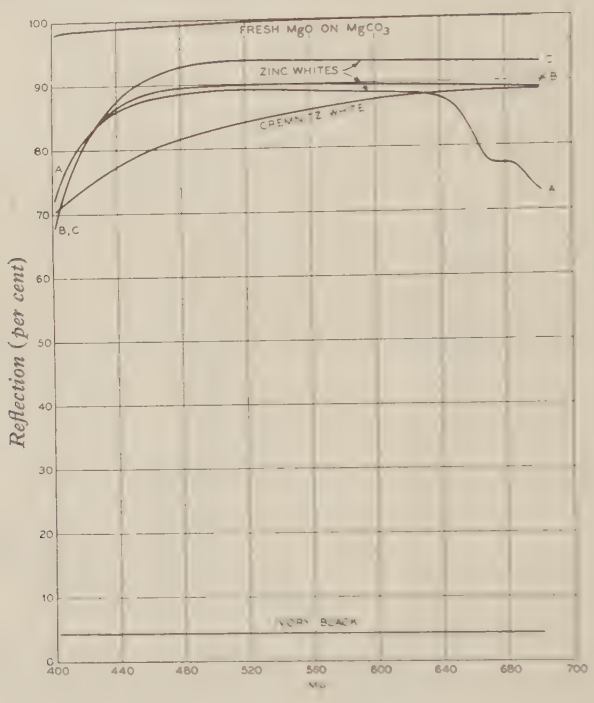


Figure 11. Reflecting power of various white pigments ground in oil.

From the standpoint of permanence, which is always questioned at once by the serious artist, the three organic pigments above listed are extraordinarily good compared with anything in this class available a decade ago. Pictures painted with them may be expected, on the basis of forced-fading tests, to last unchanged under ordinary gallery lighting for a lifetime. However, in addition to the closer approxi-

* An additional reason for the relatively poor rendering of blues is the fact that the medium (linseed oil) is definitely yellow, and, as is well known, becomes yellower with age. A decided drop in reflecting power, amounting to about 25 per cent in the extreme blue, is shown by all the spectro-photometric curves without exception. This is illustrated by the curves for several samples of zinc white, shown against magnesium oxide as standard in figure 11. The satisfactory realization of the three-colour method calls, even more than ordinary oil painting, for a colourless non-yellowing medium. Experiments which have been described on the use of silicon ester, by King, *J. Oil and Colour Chemists Assn.* (v), **13**, 28 (1930), are of great interest in this connexion.

mation to the specified spectral characteristics which is still to be attained, an even higher degree of permanence is desirable. In searching for satisfactory pigments of greater permanence I have been able to make substitutes for the yellow and the blue-green from available artists' colours which are not only more permanent but are on the whole somewhat better in performance. For the yellow I find that "extra pale cadmium yellow" (cadmium sulphide prepared at high temperature) is an exceedingly close match to the Hansa yellow 5-G both visually and spectro-photometrically. Its absorption, figure 9, extends slightly further into the green, thus contributing better toward filling the space between blue and green, while when diluted with white it follows practically the same course as does the Hansa yellow.

For the blue-green I find that the purest form of iron blue or Prussian blue, potassium ferric ferrocyanide, called by artists' colourmen "Chinese blue," is very satisfactory. In integral colour it appears too blue, but when it is examined spectrophotometrically (figure 7) the reflecting power is found to be not only higher in the deep blue but also higher in the green than the Glacier blue lake, and in addition the pigment is free from the band of reflection in the deep red. In spite, therefore, of its too blue integral colour and low reflecting power in the green, it is a nearer approximation to a minus red than is the better-appearing blue-green lake first chosen.

With this palette—extra pale cadmium yellow, Chinese blue, and a phosphomolybdo-tungstic acid lake of Rhodamine 6-G—together with zinc white as the base, I have done a great deal of painting, and some has been done by several artists whom I have been able to interest in the problem. It proves to be adequate to handle the vast majority of colours needed for naturalistic painting. In combination these colours give an excellent black; in mixtures two by two they give good reds and greens. The blues, because of the deficient reflection of both the Chinese blue and the Rhodamine lake, are not as bright as the best cobalt or ultramarine blues. A condition which makes these defects of relatively small importance is that practically all colours of natural objects are extremely unsaturated. It is practically always necessary to use some admixture of white to match the colours one desires to copy.

A few words may be said at this point in regard to the technique of three-colour painting in oils by the use of the pigments just described. The outstanding difference between the ordinary oil technique and this special one is that in place of a dozen or more pigments being placed on the palette there are only four, namely the white, which is to be made the subject of absorption, and the three subtractive primaries. It is important to keep these colours clean and free from contamination by the others. My own practice has been to use a set of four small palette knives to transfer the pigments from their original piles to a portion of the palette where they are mixed, the mixing then being done either with another knife or with the brush. Another procedure is to lay out the special colours in a series of quite small piles which can be picked up one by one with a brush or palette knife without disturbing the others. In general, because of the great saturation of these colours and their wide difference in character from any which occur in the final picture, a considerable amount of mixing is necessary before the actual colour laid on the canvas is attained. The technique which has just been described differs considerably

from the practice of many artists who form their mixtures as they paint, with the brush, which makes possible great speed of attack. This practice is aided by the comparative impurity of the ordinary artist's colours, and the consequent rather small amount of mixing which is called for. The procedure best fitted for the use of the three-colour palette is still in a state of flux, but experience thus far has indicated that its special requirements can be readily adapted to effective painting with a great shortening of the process of learning the method of getting the desired colours.

§ 5. DISCUSSION

I conclude my presentation of the theory and experimental realization of a method of trichromatic painting in oil* by a discussion of the researches necessary to its full perfection, and by a discussion of the advantages and merits of the method.

The future technical work needed follows from the comments already made on the pigments now available. Taking these up in turn, we note that the yellow at present used, namely extra pale cadmium yellow, is very satisfactory, and is probably as close an approximation to the ideal as may reasonably be expected. A defect is that it is an opaque colour, while the other two are transparent, which results in the blacks being unbalanced and going purple instead of grey on dilution with white. A possible improvement would be the substitution of a transparent yellow of high colouring-power. Turning next to the minus green, for which a tungstic acid lake of Rhodamine 6-G is used, we note two deficiencies. First, the reflecting power for the blue of the spectrum is too low, a defect which is accentuated as the concentration is increased; this pigment in fact becomes red with concentration instead of remaining crimson. Secondly, while remarkably permanent for colours of this nature, it is nevertheless subject to slow change on exposure to light. A principal problem for research is therefore the production of a crimson having the general characteristics of this one, but of higher blue reflecting power and of greater permanence†.

* Although my own painting experiments have been confined to oil colours, I have investigated water colours to the extent of looking over the water-colour paints put out by several manufacturers. Of these a pale cadmium and a Prussian blue are practically identical with the oil pigments I have used. The Rhodamine 6-G lake is obtainable as a water paste, which has very similar working properties to the water colours sold in tubes. Experiments on mixing these show that excellent matches can be made with all the colours usually included in a water-colour box. The substitution of water for the yellow oil medium results in a notable improvement over the oil mixtures in the blues. In working with aquarelle these three colours are all that are needed, as the white is that of the paper used.

† The best samples of Rhodamine 6-G lake compare favourably in permanence with the madder lakes, very widely used and approved by painters. According to Lauric these "fade slowly when exposed to light, but are sufficiently permanent for all practical purposes." Quite marked differences in fastness to light are found in lakes of the same composition, but from different sources. This fact offers promise that further research into methods of manufacture will yield material improvements in permanence. Such research, directed intensively to the few pigments called for by the three-colour palette, has greater chance of success than a diffuse attack on the whole problem of the permanence of dyestuff lakes. A possible help in improving the permanence of pigments of this class is the admixture of a considerable fraction of some permanent pigment of approximately the same colour. For instance, cobalt violet, figure 12, is suitable for using with the Rhodamine lakes. Alone it is not sufficiently saturated but in the mixture it carries part of the colour and acts as a partial light shield to the lake.

Of the three pigments the minus red is furthest from the ideal and constitutes the most difficult problem. In the dyestuff field an apparently universal characteristic of the blue-greens is that they derive their colour from two absorption bands, one in the yellow and another in the violet. They thus exhibit two reflecting regions, one in the deep red and one in the blue-green of the spectrum, and this latter band is too narrow for our purpose. Both the spectrum blue and the spectrum green, which should be present undiminished, are deficient. Consequently the blue obtained by mixing such a pigment with crimson is too dull, as compared for instance with ultramarine, and the greens obtained by mixing with yellows are not so brilliant as, for instance, emerald green. In the field of inorganic colouring materials I have not as yet been able to find a satisfactory pigment. A promising line of attack appears to lie in the copper compounds. Thus the beautiful blue-green obtained by colouring the proper kind of glass with copper, of which transmission curves are shown in figure 7, would be highly satisfactory if it could be secured in a form suitable for grinding with linseed oil. As yet, however, it has not proved possible to colour a glass nearly deeply enough to make it a good pigment when powdered. What is needed here is a compound similar to cobalt blue in which the colouring material is copper instead of cobalt, or a compound similar to the iron blues but coloured with copper. This is a problem which I recommend for study to colour chemists.

Before we leave the subject of pigments for three-colour painting a few words may be said on the subject of variations or compromises on the simple three-colour palette which might be worthy of consideration. These are of two classes; first, such compromises as may be forced by an impasse in the effort to meet the specifications indicated by theory, and secondly, such as may simplify the practical use of the method. Under the first class I may mention the possibility of using more than one pigment for each of the minus colours. The object of this would be to prevent those losses of luminosity and saturation which are due, as discussed above, both to the use of a continuous-spectrum white and to the characteristics of available pigment absorption bands. The pigment for which a substitution of two or more is most likely to be advantageous is the minus green. The Rhodamine 6-G lake, as shown in figure 8, falls off badly in reflecting power on concentration at the red and even more at the blue end of the spectrum, necessitating a lowered key for the whole painting where bright reds and still more where blues occur. The suggested compromise would be to substitute for the single pigment a pair, "warm" and "cold," one with its absorption band more toward the blue the other toward the red. In this case the pair might consist of the two Rhodamines, 6-G and B, if these can be produced of sufficient permanence. Figure 12, giving spectro-photometric reflecting-power curves for these two pigments, illustrates this point*. Another way

* The importance of the minus greens in subtractive painting is probably not generally realized by artists. These colours do not often occur as such except in flowers, and are very poorly represented in the traditional pigments. Their use in mixtures to give greys, browns and blues is not obvious from their very different appearance from these colours. Their importance will be understood upon outlining the characteristics of a full subtractive palette drafted with no reference to the simplifications made possible by the three-colour principle. Such a palette, illustrated in figure 12, would consist of a series of narrow juxtaposed intense absorption bands, which together would absorb the entire

to offset the deficiency of this crimson would be to use with it an additional minus red and minus blue which demand a lower concentration of the crimson in their mixtures. This would mean a more orange yellow, such as aurora yellow, and a bluer minus red, such as a greenish ultramarine. A complete duplicate three-colour palette might consist of strontian yellow and aurora yellow for the minus blues, Rhodamine B and 6-G lakes for the minus greens, and two minus reds, one a greenish ultramarine, the other some as yet unavailable copper compound of the

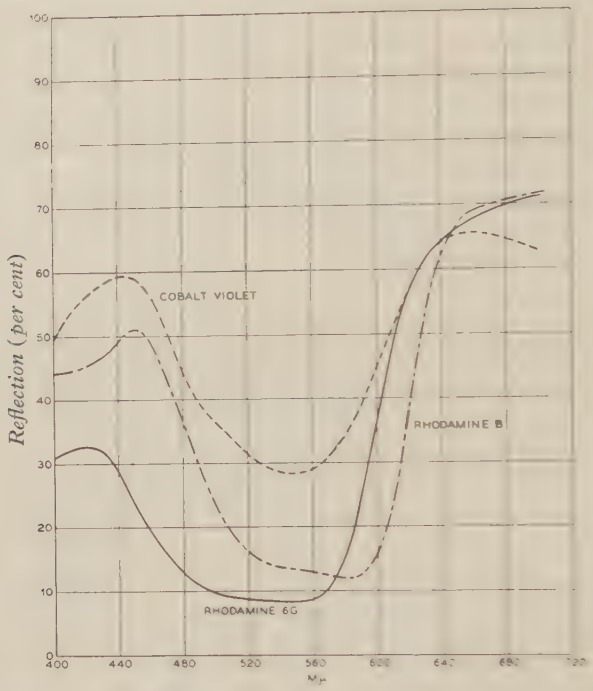


Figure 12. Reflecting power of two Rhodamine lakes, illustrating the use of two available pigments to cover the range demanded theoretically from one.

kind previously discussed. Such a duplicated palette would give an exceedingly rich set of mixed colours. It would, however, be purchased at the sacrifice of the simplicity and freedom from ambiguity of the simple three-colour palette, reasonably ideal pigments for which, it is hoped, will be ultimately available. I consider it unlikely that any greater complication than the suggested use of two minus greens need be contemplated.

The second possible variation of the three-colour palette is in practices which may simplify the technique. Of these I select the use of a series of neutral greys

spectrum. Of these the two extremes would be a lemon yellow and a light blue-green. All the others would be crimsons, purples and pinks. This complete subtractive palette may be contrasted with the "spectrum" series of pigments, put out by some colourmen, which are approximately of the principal hues of the spectrum, and obviously inspired by ideas only appropriate to additive colour mixture. A true spectrum palette of this latter sort would permit no mixtures except with white, and must be continued beyond the spectrum through a series of purples, thus demanding more pigments than the subtractive one above described; and this in turn is, as the three-colour principle shows, more extensive than necessary.

instead of white as the base from which to absorb light. By reference to the colour cube, figure 1, it is evident that instead of arriving at a colour in the cube by absorbing red, green and blue by working back from the white vertex, one could start with the proper point on the black-white or grey line, and use only two of the coloured pigments. Stated in terms of mixing technique, this means that the mixing-table or palette might display the three minus colours and a series of greys, the latter made by previously mixing white with the three coloured pigments, or by mixing white and some neutral black, such as ivory black*. The hue of the desired colour would then be obtained by mixing two of the coloured pigments, its value and saturation by choice of the proper grey. The same result is also obtained if we make

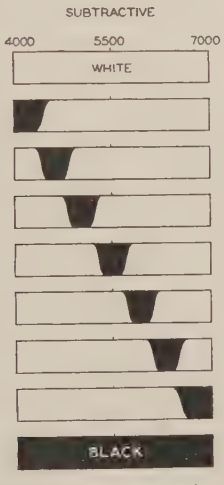


Figure 13. Characteristics of a complete subtractive palette disregarding the simplifications made possible by the three-colour principle.

our palette of the three coloured pigments, white and black. The practical simplification introduced by this procedure lies in the fact that the process of darkening a colour is such a constantly recurring and fundamental problem as to deserve a simple solution. Where it is done by adding all three minus pigments, the hue is apt to be thrown violently this way and that before the apparently simple change is made. The use of black is thus not a theoretical requirement but a working simplification comparable to the use of vectors as contrasted with Cartesian co-ordinates.

In conclusion I wish to stress the practical advantages and the philosophical merits of the method of painting I am advocating. The great advantage of the three-colour system is the reduction in the number of pigments necessary, and the simplification of the procedure of mixing colours. The mere reduction in the number of pigments used diminishes the likelihood of alteration with the passage of time in the pictures painted. It has long been commented upon by students of old paintings

* The practical advantage of using a simple black lies in the fact, already noted, that the three pigments now available are of different transparency, so that the black obtained by their mixture does not remain neutral on simple dilution with white. In making a series of greys the proportions of the pigments have to be continuously altered.

that many world-famous masterpieces have suffered serious alteration in appearance because the numerous pigments used in them have changed at different rates and directions owing to their chemical interaction and the effects of light. Spots of reflected light painted with an earth colour become, for instance, brighter than the primary source painted in a lake. Pictures noted originally for brilliant flesh tints are now admired for their monotone modelling. With the number of pigments limited to the irreducible minimum and research concentrated on the qualities of these, the likelihood of such changes should be notably reduced. As to the procedure in mixing pigments, whereas with the artists' pigments now in use the number of ways of arriving at a given colour (provided that it can be made) is enormous, with the three-colour palette there is just one way to mix each colour, and with a perfected three-colour palette there will never be any question as to whether the colour can be secured. The way to obtain a given colour is furthermore clearly indicated once the underlying theory is grasped.

This latter point leads to what is perhaps the most important philosophical merit of the three-colour palette. This is that by its use the theory of the mixture of pigments, and the education of artists in colour, are directly correlated with the theory of colour as understood by scientific students of optics. It may be safely said that at present there is no such correlation; that the theory of colour as produced by light is commonly not understood by artists, and is widely believed to be unconnected with the mixture of pigments. This latter belief is, of course, erroneous but is quite comprehensible as long as artists work with pigments which have no scientific basis.

The teaching of colour to artists need not be extended to the complete theory as here developed. I find that a quite sufficient guide to the use of the three colours is obtained by picturing the spectrum on a colour triangle, such as figure 2 *b*, with white at the centre, the third side giving the purples, and indicating by three arrows the positions of the three pigments at the sides of the triangle. A student needs merely to be told to pick out the spectrum hue of the colour he wishes to paint, and then the colours which must be mixed to produce that hue are those whose arrows lie nearest. The ability to pick the spectrum hue of a colour is quickly acquired. Tints are then made by admixture with white, and general darkening of the colour is produced by an admixture of all three or black. Devices such as superposable transparent wedges of the three minus colours also suggest themselves as possible means by which the artist of the future may be led swiftly and surely to the mastery of the purely mechanical phase of his craft.

As a final word, let me emphasize that what I have here presented is in the nature of a preliminary survey of the problem. It represents spare-time study extending over about a year, and as such cannot pretend to any finality. New pigments and new techniques in painting must be exhaustively tested by all the resources of the chemist, and particularly by the passage of time. I believe, however, that the investigation has been carried sufficiently far to show that the three-colour principle of Thomas Young is capable of materially simplifying and advancing the painter's craft. It is my earnest hope that the subject may be taken up seriously by some properly equipped research body, with the backing of some academy or other institution of art.

621.317.725

THE RECTIFYING PEAK VOLTMETER AS A STANDARD INSTRUMENT

By A. T. STARR, M.A., B.Sc. (Faraday House, London)

Communicated by Dr A. Russell, F.R.S., July 8, 1933. *Read October* 20, 1933

ABSTRACT. The paper describes a peak voltmeter using overbiased rectifying valves. Various errors are discussed and calculated.

§ 1. INTRODUCTION

The rectifying peak voltmeter has been described by Chubb and Fortescue*. The voltage is applied across a condenser *C* and a pair of rectifiers, as shown in figure 1. The rectifiers are a pair of contacts which make and break simultaneously, so that only one of the pair is closed at any moment. An ammeter *A* is placed in one arm and measures the current that passes through one rectifier. If the wave is periodic and vanishes at equal time intervals, it is possible to time the make and break so that the positive charging current (half the sine wave) will go through rectifier 1 and the negative through rectifier 2. For this method to be possible it is necessary that there should be no even harmonics, or else the wave will not vanish at equal intervals of time, the duration of the positive part of the wave being different from that of the negative.

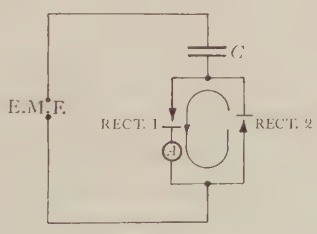


Figure 1.

In order to overcome this awkward restriction the mechanical contacts have been replaced by thermionic diodes, but the following errors are immediately introduced. There is an error due to the fact that diodes will give an emission current of about 2 per cent of their normal value, even when no voltage is applied to the anode, so that even when no e.m.f. is applied to the system in figure 1 a circulating current will flow in the direction shown. Moreover diodes have finite and variable

* *Trans. Am.I.E.E.* **32**, 739 (1913).

impedances. Corrections can be calculated* but the calculation is irksome and the correction depends upon the valve characteristic; indeed it is difficult to say of what order the error is likely to be for various applied voltages and diode characteristics, especially as the characteristics vary with time. It is possible to suppress the circulating current by applying bias batteries *B*†, as shown in figure 2. Thus, if curve 1 in figure 3 shows the (emission current, anode volts) characteristic for each unbiased diode, curve 2 is the characteristic for a suitable bias; and if this curve were a straight line and fixed in time the method would be satisfactory. If

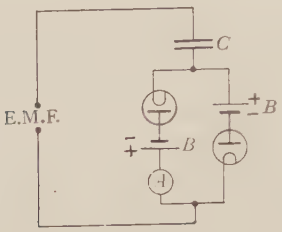


Figure 2.

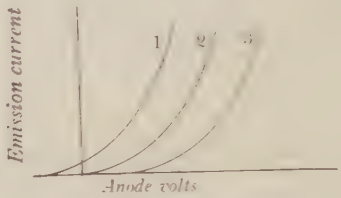


Figure 3.

the bias is excessive, on the other hand, the characteristic for each diode is shown by curve 3. It has generally been assumed that this latter state is undesirable, but it will be shown that, far from being undesirable, it is the best arrangement. It reduces the errors due to the diode impedances to a very small amount which is easily calculated within narrow limits, and, although the characteristics and biases may vary with time, the errors will always remain less than some preassigned negligible quantity. The instrument may, in fact, under these conditions be used as a standard without the necessity of continual recalibration.

The arrangement is portable and robust.

§ 2. THE MODIFIED PEAK VOLTMETER

The arrangement is shown in figure 4. The high voltage is applied across the standard air condenser and the rectifying system (shown in a rectangle) in series. The valves are triodes, Marconi-Osram DL, d.c., indirectly heated, and require 0·25 A. at 16 V. They were chosen because of their good insulation between heater and cathode, since the cathodes may have a voltage-difference of 100 V. and the heaters are nearly at the same voltage. The valves are heated with a.c. at 50 ~ at 230 V., transformed down to approximately 16 V. by a Heayberd transformer with a ratio of 250:18. The small resistance r is inserted to produce the correct current in the heaters. A is a standard micro- or milliammeter. A safety (neon) gap is placed across the device in case of a failure.

* R. Davis, G. W. Bowdler and W. G. Standring, *J.Inst.E.E.* **68**, 1228 (1930).

† See *Methods of High Voltage Measurement* by S. Whitehead, L. T. 55, British Electrical and Allied Industries Research Association, p. 47.

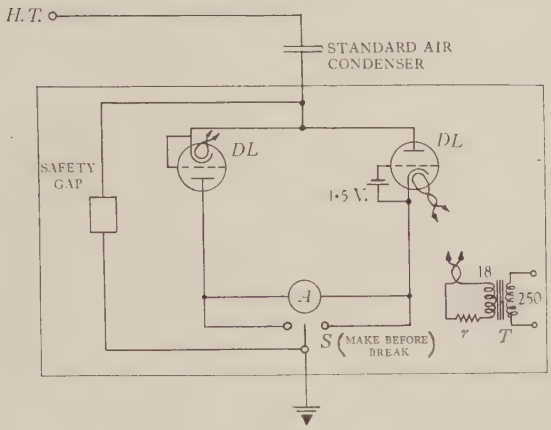


Figure 4.

In one valve the grid is connected to the cathode, whilst in the other a grid-bias of about $-$ 1·5 V. is used in order to suppress the circulation current. Table 1 gives the values of circulating current for different values of the grid-bias in this valve.

Table 1.

Grid-bias (V.) negative	0	0·2	0·4	0·6	0·8	1·0	1·2	1·5
Circulating current (μA.)	280	230	110	56	22	7	0	0

It is seen that the correct bias for the suppression of the circulating current is 1·1 V., but it will vary a little from valve to valve. It will be shown in appendix 1 that overbiasing produces no appreciable error in the instrument. A variable grid-bias is cumbersome and quite unnecessary, but it is easy to put a small dry cell of 1·5 V. between the grid and the cathode.

Assuming that the biasing is perfect and the diode impedances zero, the simplified theory of the instrument is as follows. The voltage to be measured charges up the capacity C. When the voltage decreases continuously from V_{max} to $-V_{max}$, the charging current goes down through the left-hand diode and the total charge that flows is $2CV_{max}$. If the fundamental frequency of the wave is f, this happens f times per second, so that the mean current i_g in the galvanometer is given by

C, V_{max}

f

$$i_g = 2fCV_{max} \qquad \text{.....(1).}$$

i_g

Therefore

$$\left.\begin{aligned} V_{rms} &= V_{max}/\sqrt{2} \\ &= i_g/2\sqrt{2}fC \\ &= (\pi/\sqrt{2})\, i_g/\omega C \\ &= 2{\cdot}22\, i_g/\omega C \end{aligned}\right\} \qquad \text{.....(2),}$$

where

$$\omega = 2\pi f.$$

ω

Equation (1) holds for any wave, provided it vanishes at equal time intervals and has only one maximum and one minimum per cycle. Equation (2) holds only for a pure sine wave, since it assumes that $V_{rms} = V_{max}/\sqrt{2}$.

§ 3. ERROR DUE TO OVERBIASING

e Suppose curve 1 in figure 5 is the characteristic for one diode and curve 2 for the other, where these curves are straight lines that meet the axis of voltage at $-e$ and $-e$. In practice the curves will not be straight lines, but the deviation will be considered in a moment and shown to be negligible. The combination of the diodes i acts like a constant resistance, which requires a flash-over voltage. If i is the current v and v the voltage across the rectifiers,

R
$$v = e + Ri, \text{ for } i > 0 \quad \text{.....(3).}$$
and
$$v = -e + Ri, \text{ for } i < 0$$
For
$$-e \leqslant v \leqslant e, \qquad i = 0 \quad \text{.....(4).}$$

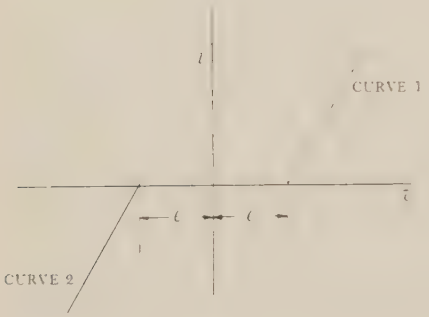


Figure 5.

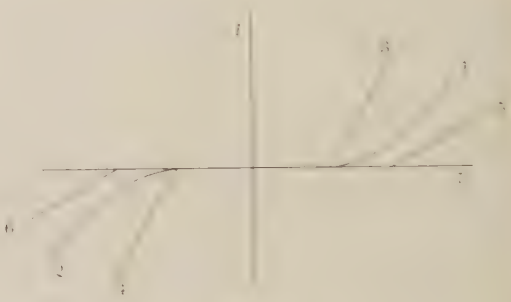


Figure 6.

It is proved in appendix 1 that the error in peak-voltage reading is given by

$$\tfrac{1}{2}(\omega CR)^2 V_{max} + e \quad \text{.....(5),}$$

so that the error is

$$[\tfrac{1}{2}(\omega CR)^2 + (e/V_{max})] \times 100 \text{ per cent} \quad \text{.....(6).}$$

If the characteristics are curves, as shown by curves 1 and 2 in figure 6, it is clear that the meter will read more than if the characteristics were 5 and 6, but less e_1, R_1 than if they were 3 and 4. If curves 3 and 4 have intercepts $-e_1$ and slope R_1 and e_2, R_2 curves 5 and 6 have intercepts $+e_2$ and slope R_2, the error in the actual case lies between

$$[\tfrac{1}{2}(\omega CR_1)^2 + (e_1/V_{max})] \times 100 \text{ per cent and } [\tfrac{1}{2}(\omega CR_2)^2 - (e_2/V_{max})] \times 100 \text{ per cent}$$
$$\text{.....(7).}$$

The values of e and R were found roughly for the practical arrangement shown in figure 4. Figure 7 is a graph of v plotted against i, and it is found that

$$e = 13 \text{ V.},$$
and
$$R = 4400\ \Omega. \quad \text{.....(8).}$$

At 50 ~ and with a condenser of 100 $\mu\mu$F., $\omega CR \approx 10^{-4}$, so that the error due to the impedance is of the order 10^{-6} per cent, which is quite negligible. The error due to overbiasing is then a constant, namely e. That is to say, that no matter what

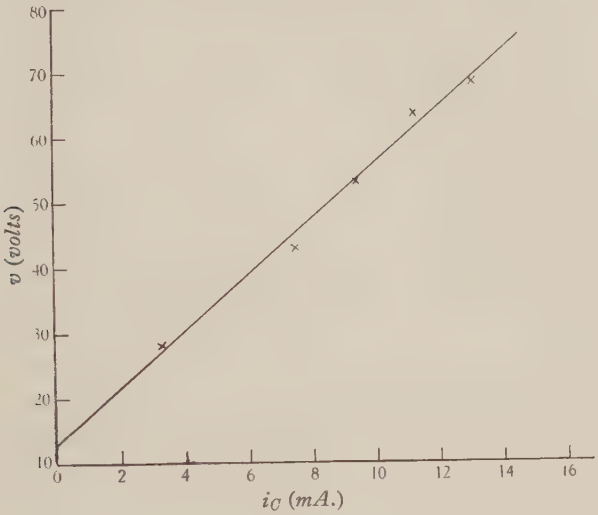


Figure 7.

voltage may be applied, the peak is underestimated by the constant e which equals 13 V. nearly. We can get the true peak by adding e to the voltage $i_g/2fC$, so that

$$V_{\text{peak}} = (i_g/2fC) + e \qquad \text{.....(9),}$$

or

$$V_{\text{rms}} = (\pi/\sqrt{2})\, i_g/\omega C + e/\sqrt{2} \qquad \text{.....(10).}$$

Tests were made with the results shown in table 2.

Table 2.

Kilovolts (r.m.s.) applied to primary transformer winding and standard voltmeter	i_g (μA.)	$2{\cdot}22\, i_g/\omega C$	$2{\cdot}22\, i_g/\omega C + e/\sqrt{2}$	Error (per cent)
1·00	34·2	982	991	− 0·9
2·00	69·5	1992	2001	0·0
3·04	105·5	3050	3059	+ 0·6
4·04	140·0	4020	4029	− 0·3
5·06	177·0	5080	5089	+ 0·6
6·10	212·0	6090	6099	0·0
7·16	248·0	7140	7149	− 0·1
8·20	282·0	8100	8109	− 1·2
9·26	320·0	9200	9209	− 0·55

$C = 246\ \mu\mu$F. for the standard condenser.

Even if the error e due to overbiasing is neglected, the error above 2 kV. is very small, about 0·4 per cent. If no biasing is used, the circulating current is

280 μA. which corresponds to an applied voltage of 8·2 kV. The overbiased arrangement thus enables comparatively low voltages of 2 and 3 kV. to be read accurately. Very high voltages will be read accurately on the under- and over-biased arrangements, but the latter extends the range right down to 2 kV. With a condenser of 246 $\mu\mu$F., the overbiased voltmeter will read accurately from 2 kV. to 500 kV., but it is not quite certain when the underbiased voltmeter will begin to be accurate. This will be investigated in the next section, but before this is done a surprising fact must be explained.

Suppose 2 kV. (r.m.s.) is applied to the voltmeter. The galvanometer current is about 70 μA. and the condenser current is 2·22 · 70 or 155 μA. (r.m.s.). This means that the voltage across the diodes is never more than 15 volts, whilst no current flows when the voltage is less than 13 volts. One would expect that the error would be of the order 1300 15 per cent or 80 per cent. The reason why the error is small is this: when the alternating e.m.f. is in series with the condenser C and resistance R, the current leads the voltage by nearly 90° if ωCR is small. The voltage across the resistance (here the rectifiers) is proportional to the current, and is a minimum when the applied e.m.f. is a maximum. This is due to the fact that the condenser is fully charged at this moment, and the potential of charge opposes the applied e.m.f. almost completely. If the rectifiers will not pass current when the voltage lies between $-e$ and $+e$, the charge and potential difference of the condenser remain constant for a time, whilst the applied e.m.f. must change by $2e$; and as the condenser charge remains unaltered, the voltage across the rectifier changes by $2e$ also in this time. It takes only a very short time for the e.m.f., $V_{\max}\cos\omega t$, to vary through a voltage $2e$ in the neighbourhood of its maximum, the time being approximately $4\omega^{-1}\sin^{-1}\sqrt{(e/2V_{\max})}$ sec. or $2\pi^{-1}\sin^{-1}\sqrt{(e/2V_{\max})}$ of a cycle. If e is small compared with $V_{\max}$, as it is, current does not flow for $\pi^{-1}\sqrt{(2e/V_{\max})}$ of a cycle. If $e = 13$ and $V_{\max} = 2000\sqrt{2} = 2820$, this is only 1/30 of a cycle, so that current is suppressed only for a short time. Also the current that should flow is small in comparison with the mean current, so that the error is very small.

§ 4. ERROR DUE TO CIRCULATING CURRENTS

In order to see the effects of circulating currents, various voltages from 20 to 90 kV. were applied with grid-bias − 1·5 and grid-bias 0. The capacity of the standard condenser was, in this case, 100 $\mu\mu$F. First it was verified that the grid-bias − 1·5 V. suppressed the circulating current entirely, and it was assumed then that this arrangement gave the correct result. The voltmeter in the primary supply was not at all accurate, but served merely as an indication of the order of the applied voltage. Then the galvanometer readings for biases of − 1·5 V. and 0 V. were found for voltages 20 to 90 kV. The error due to the circulating current was then found to be as shown in table 3.

Table 3.

Applied kilovolts (approximate)	i_g (μA.)		Error (per cent)
	Bias − 1·5	Bias 0	
20	252	296	17·5
30	416	442	6·25
40	540	559	3·5
60	822	830	0·97
80	1086	1093	0·645
90	1220	1230	0·82

It seems then that the underbiased system cannot be used for voltages less than 100 kV. without an error of nearly 1 per cent, whereas the overbiased system can be used down to 2 kV.

§ 5. ERROR DUE TO MULTIPLE MAXIMA AND MINIMA

Suppose the wave is of the shape shown in figure 9. Then the voltage that will be measured will be $(v_1 - v_2) + (v_3 - v_4) + (v_5 - v_6)$, since each of these downward steps will send charging current through the same diode. The measured voltage is thus

$$(v_3 - v_6) + (v_1 - v_2) + (v_5 - v_4),$$

instead of $(v_3 - v_6)$, which is the difference in peaks. The voltmeter will thus overestimate the peak voltage.

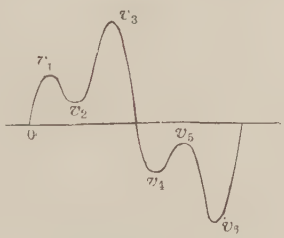


Figure 9.

There are various causes that produce multiple maxima and minima. One cause is bus-bar discharge. It was found, however, that it required a considerable artificial discharge, much in excess of what will ordinarily happen, to cause an increase of reading on the overbiased voltmeter. It is not known, but it is expected, that overbiasing helps here too. If the discharge voltage is less than e it will be ignored, whereas in the unbiased instrument any discharge voltage will have an effect which will be greater the higher the frequency of the discharge, since condenser current is proportional to frequency.

Error due to third harmonic. A third harmonic, provided it is large enough and has the requisite phase, can cause multiple maxima and minima. Phase is important, as can be shown by taking the two simple cases of figures 10 and 11. In figure 10

the fundamental is $E \sin \omega t$ and the harmonic is $E_3 \sin 3\omega t$, say. The condition for multiple maxima and minima is that

$$\frac{d}{dt}(E \sin \omega t + E_3 \sin 3\omega t) = 0$$

should have real roots apart from $\omega t = \frac{1}{2}\pi$. The equation is

$$\begin{aligned} 0 &= E \cos \omega t + 3E_3 \cos 3\omega t \\ &= \cos \omega t \,[E + 3E_3 (4 \cos^2 \omega t - 3)], \end{aligned}$$

giving

$$4 \cos^2 \omega t = 3 - E/3E_3.$$

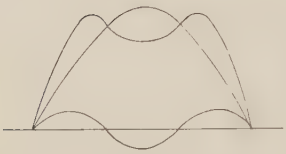


Figure 10.

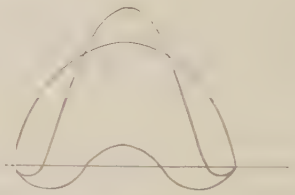


Figure 11.

For real roots,

$$3 - E/3E_3 > 0,$$

or

$$E_3 > \tfrac{1}{9}E,$$

so that multiple maxima and minima will occur for a harmonic greater than 11 per cent.

It is clear that in the case shown in figure 11 the condition for multiple maxima is that the equation

$$4 \cos^2 \omega t = 3 + E/3E_3$$

should have real roots, so that $E/3E_3 \geq 3 - 4$, and $E_3 \geq \frac{1}{9}E$; i.e., a harmonic of 33·3 per cent is required.

Suppose the fundamental is $E \sin(\omega t + \alpha)$ and the harmonic is $E_3 \sin 3\omega t$. We may call α the "phase difference" between these waves, but it must be remembered that this definition is arbitrary. To cover all possible cases it is sufficient to take α between 0 and $\pi/3$, for if α is negative, we may reverse the sign of α and obtain the wave reversed in time; if α is greater than $2\pi/3$ or $4\pi/3$ we may subtract $2\pi/3$ or $4\pi/3$, for this means starting one or two cycles later in the third-harmonic wave; and if α lies between $\pi/3$ and $2\pi/3$, we can subtract $2\pi/3$, obtaining an angle between 0 and $-\pi/3$, whose sign we may reverse. Thus in all cases we can put $0 \leq \alpha \leq \pi/3$. The wave is then

$$E \sin(\omega t + \alpha) + E_3 \sin 3\omega t \qquad \ldots\ldots(11).$$

The r.m.s. value of the wave is

$$\sqrt{(E^2 + E_3^2)} \quad \text{or} \quad E\sqrt{\{1 + (E_3/E)^2\}},$$

which is approximately $E(1 + E_3^2/2E^2)$. This is independent of α, the phase.

Figure 12 gives the ratio of peak to fundamental peak for various magnitudes and phases of third harmonic. In the cases of $\alpha = 0$ and $\alpha = 10°$ multiple maxima

occur and the rectifier reading is greater than the actual peak. Even with this effect, for $\alpha = 0$, the rectifier reading is less than the peak of the fundamental for harmonics up to 23·5 per cent.

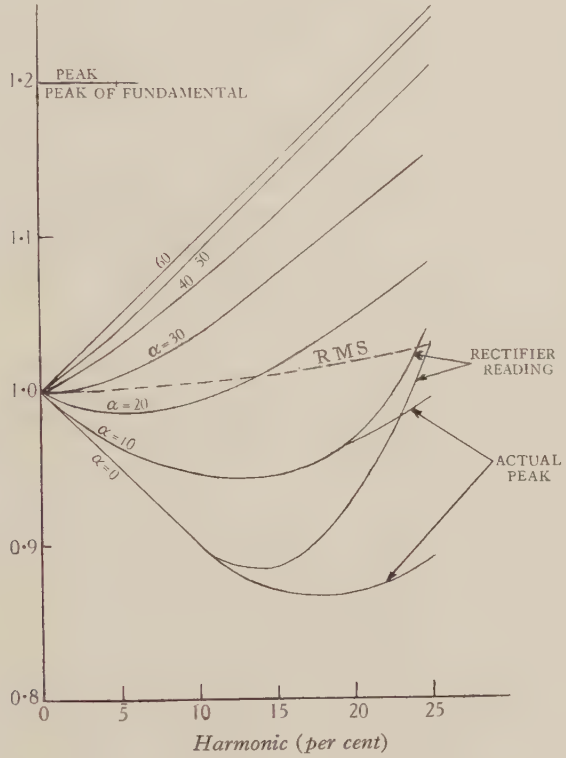


Figure 12

§ 6. ERROR DUE TO CAPACITY ACROSS RECTIFIERS

Suppose that the screened leads that connect the standard condenser C to the rectifiers have a capacity C_1 connected across the rectifiers, as shown in figure 13. This capacity will shunt some current from the rectifiers and will cause an error, which may be calculated very quickly by the use of Thevenin's theorem*. By this

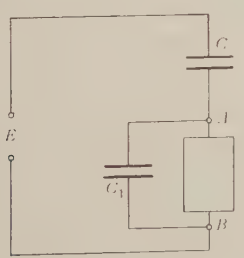


Figure 13.

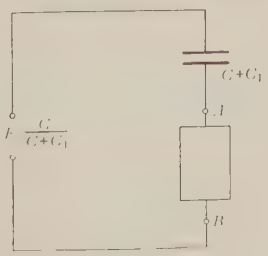


Figure 14.

* Shea, T. E., *Transmission Networks and Wave Filters*, p. 55.

theorem, we may replace the whole of the system to the left of AB by the e.m.f. across AB when the rectifier system is absent, and a series impedance equal to that looking left from AB, when the source of e.m.f. is short-circuited.

The e.m.f. across AB when the rectifiers are absent is $EC/(C - C_1)$. The impedance looking left is that of condensers C and C_1 in parallel. Thus the effect of C_1 is to reduce the e.m.f. to $EC/(C - C_1)$ and increase the capacity to $(C - C_1)$, figure 14. The charge is $[EC/(C - C_1)]$ $(C - C_1)$ or EC, so that if the rectifiers had no impedance the error would be zero. As it is, the error increases from $\frac{1}{2}(\omega CR)^2 E$ to $\frac{1}{2}(\omega . C - C_1 . R)^2 E$. The former is negligible and the latter is nearly $\frac{1}{2}(\omega C_1 R)^2 E$, so that the error is $\frac{1}{2}(\omega C_1 R)^2$ 100 per cent. With ω equal to 314 and R to 4000 Ω, if an error of $\frac{1}{2}$ per cent is permitted we can tolerate

$$(\omega C_1 R)^2 \times 100 = 1,$$

$$\omega C_1 R = \tfrac{1}{10},$$

$$C_1 = 1/10\,\omega R = \tfrac{1}{12}\ \mu\text{F. approximately.}$$

§ 7. ACKNOWLEDGMENTS

The author wishes to thank Dr L. G. Brazier for much helpful criticism and advice; and Callender's Cable and Construction Co., Ltd. for permission to publish this work, which was done in the Company's research laboratories at Wood Lane.

APPENDIX 1

Suppose that a voltage $E \sin(\omega t - \theta)$ is applied in series with a condenser C and the rectifying system. The condenser plates have charges $+q$ and $-q$, the current is i and the voltage across the rectifying system is v. Then

$$i = dq/dt.$$

The rectifiers are overbiased so that no current passes unless v is greater numerically than e.

Then
$$v = e + Ri, \text{ for } i > 0$$
$$e \leqslant v \leqslant e, \text{ for } i = 0,$$
and
$$v = -e + Ri, \text{ for } i < 0.$$

Everything is anti-symmetrical in the two half-cycles.

Figure 8 shows the charge and current as functions of time. We have the equation

$$E \sin(\omega t + \theta) = (q/C) + v \qquad \ldots\ldots(1),$$

which holds throughout, and the following additional equations which hold in the specified intervals of time.

Intervals	Equations	
(a) $t_2 - \pi/\omega < t < t_1$	$i > 0.\ v = e + Rdq/dt$	.....(2),
(b) $t_1 < t < t_2$	$i = 0.\ dq/dt = 0$	.....(3),
(c) $t_2 < t < t_1 + \pi/\omega$	$i < 0.\ v = -e + Rdq/dt$	.....(4),
(d) $t_1 + \pi/\omega < t < t_2 + \pi/\omega$	$i = 0.\ dq/dt = 0$	.....(5).

There are the following boundary conditions:

$$\text{At } t = t_2 - \pi/\omega,\ i = 0,\ v = e,\quad q = -Q \qquad \ldots\ldots(6),$$

$$t = 0,\quad q = 0 \qquad \ldots\ldots(7),$$

$$t = t_1,\quad i = 0,\ v = e,\quad q = Q \qquad \ldots\ldots(8),$$

$$t = t_2,\quad i = 0,\ v = -e,\ q = Q \qquad \ldots\ldots(9).$$

In the interval (*a*), we have from equations (1) and (2)

$$E \sin(\omega t + \theta) = (q/C) + e + R\,(dq/dt),$$

giving

$$q = A \exp.(-t/CR) - Ce + CE\,[1 + (\omega CR)^2]^{-1/2} \sin(\omega t + \theta - \alpha) \qquad \ldots\ldots(10)$$

where $\alpha = \arctan \omega CR.$

Then

$$i = -(A/CR) \exp.(-t/CR) + \omega CE\,[1 + (\omega CR)^2]^{-1/2} \cos(\omega t + \theta - \alpha) \qquad \ldots\ldots(11).$$

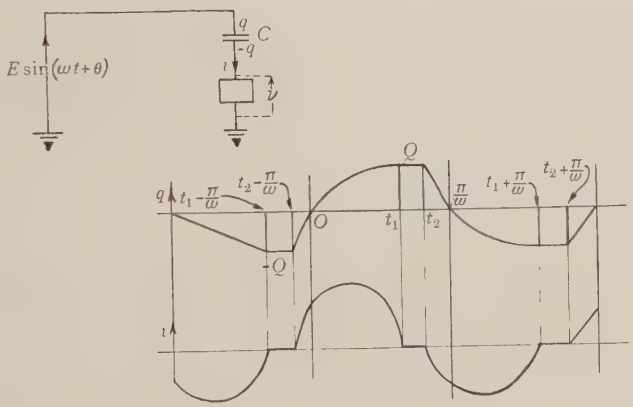


Figure 8.

In the interval (*b*) $dq/dt = 0.$

So that $q = \text{constant} = Q, \text{ say,} \qquad \ldots\ldots(12),$

and $i = 0.$

Because of the skew symmetry, we need not consider any other intervals.

Equations (6) and (11) give

$$0 = -(A/CR) \exp.[(\pi - \omega t_2)/\omega CR] + \frac{\omega CE}{\sqrt{\{1 + (\omega CR)^2\}}} \cos[(\omega t_2 + \theta - \alpha - \pi)].$$

Equations (8) and (11) give

$$0 = -(A/CR) \exp.[-\omega t_1/\omega CR] + \frac{\omega CE}{\sqrt{\{1 + (\omega CR)^2\}}} \cos[\omega t_1 + \theta - \alpha].$$

By division, these last two equations give

$$\exp.[(\pi + \omega t_1 - \omega t_2)/\omega CR] = \cos(\omega t_2 + \theta - \alpha - \pi)/\cos(\omega t_1 + \theta - \alpha) \qquad \ldots\ldots(13).$$

Equation (6) gives also

$$e = E \sin(\omega t_2 - \pi + \theta) - (-Q)/C,$$

or

$$e = -E \sin(\omega t_2 + \theta) + Q/C \qquad \ldots\ldots(14).$$

Equation (8) gives

$$e = E \sin(\omega t_1 + \theta) - Q/C \qquad \ldots\ldots(15).$$

Equations (13), (14) and (15) determine t_1, t_2 and Q, which are the quantities of interest.

Let
$$\left.\begin{aligned}\beta &= \tfrac{1}{2}\pi - (\omega t_1 + \theta - \alpha)\\ \text{and}\quad \gamma &= (\omega t_2 + \theta - \alpha) - \tfrac{1}{2}\pi\end{aligned}\right\} \qquad \text{.....(16).}$$

Then equations (13), (14) and (15) become

$$\exp.\,[(\pi - \beta - \gamma)/\omega CR] = \sin\gamma/\sin\beta \qquad \text{.....(13}^1\text{),}$$

$$2e/E = \cos(\beta - \alpha) - \cos(\gamma + \alpha) \qquad \text{.....(17),}$$

and
$$\frac{2Q}{CE} = \cos(\beta - \alpha) + \cos(\gamma + \alpha) \qquad \text{.....(18).}$$

Equations (13^1) and (17) give β and γ, and then (18) gives Q. If the system is correctly biased $e = 0$, and

$$\beta = \gamma = 0.$$

If the system is overbiased, β and γ are positive and small. $\pi - \beta - \gamma$ = 0, and in fact $\fallingdotseq \pi$. Then exp. $(\pi/\omega CR)$ is very large, since ωCR is of the order of 10,000. Therefore β is very small; in fact we may put it equal to zero.

Equation (17) then gives

$$\cos(\gamma + \alpha) = \cos\alpha - 2e/E,$$

and equation (18) gives

$$\begin{aligned}(2Q/CE) &= \cos\alpha + \cos(\gamma + \alpha)\\ &= 2(\cos\alpha - e/E).\end{aligned}$$

If the rectifier is ideal

$$Q = CE.$$

Therefore
$$Q/Q_{\text{ideal}} = \cos\alpha - e/E.$$

Since
$$\alpha = \arctan\omega CR,$$

$$\begin{aligned}\cos\alpha &= 1/[1 + (\omega CR)^2]^{1/2}\\ &\fallingdotseq 1 - \tfrac{1}{2}(\omega CR)^2.\end{aligned}$$

Therefore
$$Q/Q_{\text{ideal}} = 1 - \tfrac{1}{2}(\omega CR)^2 - (e/E).$$

The error ratio is thus

$$\tfrac{1}{2}(\omega CR)^2 + (e/E) \text{ or } \delta, \text{ say.}$$

The charge $2Q$ passes through the galvanometer once in each cycle, so that

$$\begin{aligned}i_g &= 2fQ\\ &= 2fCE(1 - \delta),\end{aligned}$$

and the maximum voltage
$$E = i_g(1 + \delta)/2fC.$$

A simpler way of expressing the voltage is

$$E = (i_g/2fC) + e,$$

i.e. the peak voltage is measured e volts low.

If the wave is sinusoidal, the r.m.s. voltage is measured $e \sqrt{2}$ volts low.

519.24

GRAPHIC STATISTICS

By A. F. DUFTON, M.A., D.I.C.

Received June 1, 1933. *Read October* 20, 1933.

ABSTRACT. The plotting of frequency-distributions is discussed. In comparing for different populations the frequency-distributions of a particular variate it is sometimes convenient to take one population as standard and to represent its distribution by a straight line. It is pointed out that the method of plotting individual points described by Hazen is incorrect.

§ 1. PERMILLE PAPER

In 1914, in a paper on "Storage to be provided in impounding reservoirs for municipal water supply*," Mr Allen Hazen described an elegant method of determining the probability of an event by plotting statistics upon special graph paper†. The method is not very widely known, and in 1930, in ignorance of Hazen's work, I put forward the same method of graphically representing a frequency-distribution. It has the advantage of showing at a glance whether the distribution is normal, and I ventured to give the name *permille* to the co-ordinate paper upon which the curves are plotted‡.

The observations determining a frequency-distribution are usually divided into classes of equal interval, and in this method of representation the figures are added up so as to give a total frequency not exceeding the upper limit of each class interval. Ordinates are then erected on a horizontal base to represent, to a special scale, these integrated frequencies as parts per mille; and a smooth curve may be drawn through the tops of the ordinates. The special frequency scale is the scale of deviates of the normal curve for each permille of frequency§. When drawn to this scale a normal frequency-distribution is of course represented by a straight line.

§ 2. PLOTTING INDIVIDUAL VALUES

With a limited number of observations, just as it is more expedient to find the median by inspection without determining classes of equal interval, so also it is more precise, and sometimes easier, to plot a frequency curve directly from the observed values. Galton's ogive was drawn in this manner||. Furthermore, whatever the number of observations, it is always desirable to plot the outlying values as individual points.

* *Trans. Amer. Soc. Civil Engineers*, **77**, 1539 (1914).

† Procurable under the name *Arithmetic Probability Paper*, from Codex Book Co. Inc., Norwood, Massachusetts.

‡ *Phil. Mag.* **10**, 566 (1930).

§ *Tables for Statisticians and Biometricians*, p. 1. Printed by Cambridge University Press and issued by the Biometric Laboratory, University College, London.

|| *Phil. Mag.* **49**, 33 (1875).

With $(n-1)$ *observations, the frequency curve can be drawn by plotting the* r*th value in order of magnitude against the ratio* r/n*.* This can be shown by Dedekind section, from the inequalities $r/(n-1) > r/n > (r-1)/(n-1)$; for if, for instance, the rth value were plotted against a ratio q less than r/n, it would be possible to find a ratio p, where $q < p < r/n$, which would divide the observations in a ratio greater than or equal to $r/(n-1)$, and it would follow that $p \geq r/(n-1) > r/n$, which is absurd.

With nine observations, the values in order of magnitude are plotted against 100, 200, 300, ... 900 per mille. When ninety-nine observations are available, it is generally convenient to plot (on permille paper) every tenth value in order of magnitude together with the five values at each end of the range. This is illustrated in figure 1, which shows a frequency-distribution of the strength of cowhair.

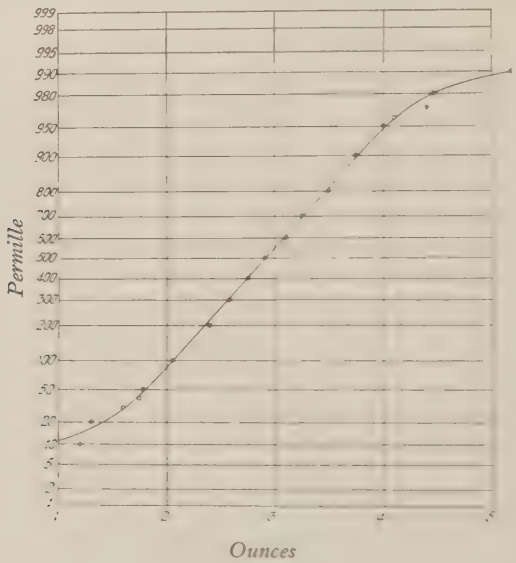


Figure 1.

§ 3. COMPARISON WITH A STANDARD POPULATION

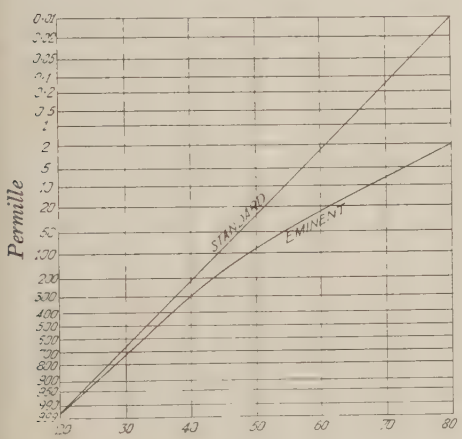
Now it is sometimes necessary to compare for different populations the frequency-distributions of a particular variate and, for this purpose, when the distributions are not normal, a different representation may be convenient, particularly when one population can be regarded as standard.

An attempt to show that capability is in some degree an inheritable character acquired by those of mature years involved the comparison of the frequency-distribution of the paternal age (at birth) for 1000 eminent persons with that for 100,000 children. The first comparison was conveniently made by means of permille

* Mr Hazen falls into error in stating (*loc. cit.* p. 1549) that, if there are 50 terms in the series, the first will be plotted on the 1 per cent line, the second on the 3 per cent line, etc.

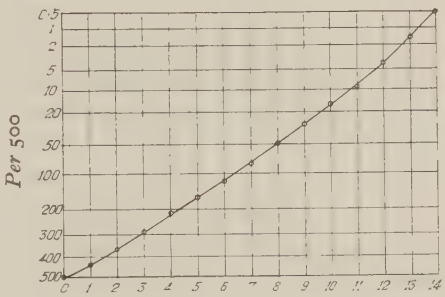
paper*, but when the possibility was considered of including in the same diagram the frequency-distributions of the paternal age for other groups of men it became apparent that the comparison would be more apt if the figure were anamorphosed so that the standard curve became a straight line; the comparison is now not with a theoretical distribution but with an actual one.

Such an anamorphosis was, indeed, suggested in my previous paper, the proposal then being to modify the scale of the variate. In general, however, the functional scale so obtained is not simple and the result is not so elegant as that obtained by modification of the frequency scale. The resulting graph is shown in figure 2 and it may be of interest to observe that in this particular example the scale of frequency is substantially logarithmic over the range from 0·01 to 100 per mille.



Paternal age

Figure 2.



Birth order

Figure 3.

§ 4. CHOICE OF METHODS

The choice of the various methods of representing a frequency-distribution must of course be made with reference to all the considerations peculiar to a particular case. The distribution of the birth order of children, of which a representation was recently required, affords an illustration of this. A first attempt, made by plotting the individual frequencies upon semi-logarithmic paper, gave quite a good graph, the points appearing to lie upon a parabola. The distribution evidently approximated to one half of a normal frequency-distribution, and I realized that the data might be represented more elegantly upon half a sheet of permille paper as in figure 3, which shows the proportion of children whose birth ordinal exceeds a specified integer.

* *Nature*, **130**, 508 (1932).

DISCUSSION

Dr L. E. C. Hughes. The curves exhibited by the author are similar to those obtained when the distribution of marks in examinations is plotted. The candidates are placed in order along the base of the diagram and given ordinates representing the percentage marks. In marking candidates' work, particularly when different groups of papers are weighted against each other and candidates are competing but with different subjects, there is always difficulty in taking into account the fact that the candidates may be faced with papers of unequal difficulty, which are subsequently marked to different standards. Double markers consistently differ by 10 or even 20 per cent. It would appear that a standard method of marking ought to be put forward with sufficient authority to ensure general adoption and capable of eliminating the two difficulties mentioned, viz. variation in stiffness of papers as far as the candidates are concerned, and variation in the standard of marking. The scheme whereby the marks are altered by a factor which brings as many above the 50 per cent line as there are below and thereby unifies the standard of stiffness and of the marking before the papers are weighted or the class percentages are declared, has much to recommend it, and the author's views on this matter would be of value.

Logarithmic paper has long been in use by acoustic engineers to represent the responses of parts of, and complete, sound-reproducing systems, as it is found that departures from ideal criteria in sound-reproduction are more correctly represented by a logarithmic scale of frequency and logarithmic scales of power or amplitude. Response curves so drawn give approximately the right weighting of changes of frequency and power-level on the ear.

Author's reply. I have no personal experience in the marking of examination papers but I believe that it is now not unusual for papers to be marked in such a fashion that about forty per cent of the candidates score between forty and sixty marks out of a hundred and that the frequency-distribution of marks is substantially normal. An examiner of experience should be able to set a paper and prepare a schedule of marks so as to secure such a distribution. A chief examiner must, of course, standardize the work of assistant examiners marking the same paper.

In a paper read last September before a joint session of Sections J and L of the British Association, Prof. Drever pointed out that an examination paper may be differently assessed not only by different examiners but also by the same examiner at different times.

535·338: 546.431

THE BAND SPECTRUM OF BARIUM OXIDE

By P. C. MAHANTI, M.Sc.

Communicated by Prof. P. N. Ghosh, July 3, 1933, *and in revised form September* 5, 1933. *Read November* 3, 1933.

ABSTRACT. New measurements of the bands of barium oxide have been made from moderate-dispersion and high-dispersion spectrograms. The band-head data of the early investigations have been extended in the red region as far as λ 8000. The bands lying between λ 8000–λ 4300 have been assigned to a single system. The equation representing the wave-numbers of band-heads is

$$\nu = 16815{\cdot}89 + \{495{\cdot}36\,(v' + \tfrac{1}{2}) - 1{\cdot}12\,(v' + \tfrac{1}{2})^2\} - \{671{\cdot}48\,(v'' + \tfrac{1}{2}) - 2{\cdot}20\,(v'' + \tfrac{1}{2})^2\}.$$

The bands (0, 3), (0, 2), (1, 2), (1, 1) and (2, 1) have been analysed. Their rotational structure reveals that the band system is due to a $^1\Sigma \rightarrow {}^1\Sigma$ transition. The following rotational constants have been evaluated:

$B_e' = 0{\cdot}30995$ cm.$^{-1}$	$B_e'' = 0{\cdot}36440$ cm.$^{-1}$
$\alpha' = 0{\cdot}0011$	$\alpha'' = 0{\cdot}0016$
$D_e' = -0{\cdot}483 \times 10^{-6}$	$D_e'' = -0{\cdot}430 \times 10^{-6}$
$\beta' = 0{\cdot}24 \times 10^{-9}$	$\beta'' = -1{\cdot}2 \times 10^{-9}$
$r_e' = 1{\cdot}938 \times 10^{-8}$ cm.	$r_e'' = 1{\cdot}787 \times 10^{-8}$ cm.
$I_e' = 89{\cdot}44 \times 10^{-40}$ gm. cm.^{2}	$I_e'' = 76{\cdot}04 \times 10^{-40}$ gm. cm.^{2}

§ 1. INTRODUCTION

When barium salts are introduced into a carbon arc in air, the spectrum shows a large number of bands consisting of several sequences which degrade towards longer wave-lengths and extend from the infra-red to the ultra-violet regions. In Kayser's *Handbuch der Spektroscopie*, volumes 5 and 7, one finds a record of the bands in the visible region as found by the early investigators when different compounds of barium were subjected to a variety of conditions of excitation, but there had been a great deal of speculation about the identity of the emitter. A systematic investigation of the spectra of barium halides in oxy-coal gas flame was conducted by Olmsted[1]. He measured the bands in the region λ 5865–λ 3810 and classified them into two groups. The group of narrow bands occupying the region λ 4890–λ 4850 was attributed to the oxide while the second group, consisting of the remaining bands, was associated with the metal. This view was later supported by Börsch[2], who extended the measurements of Olmsted in the red region up to λ 7100. But it was in conflict with the observations of Hartley[3], who noticed that the bands of the second group would appear only when the metal was just beginning the process of oxidation, and was thus led to

believe that they are due to the oxide. Eder and Valenta[4] were also of the same opinion and measured the band-heads as far as λ 3345. But it was Harnack[5] who definitely settled the identity of their emitter. He extended the measurements of Eder and Valenta further in the ultra-violet up to λ 2300, and observed that these bands disappeared in an atmosphere of hydrogen and that the presence of oxygen was essential for their production. Their rotational analysis reveals that they are due to a $^1\Sigma \rightarrow {}^1\Sigma$ transition. The odd multiplicity requires that the emitting molecule must have an even number of electrons. This definitely excludes the ionized-oxide molecule as the probable emitter, and the only possible conclusion is that the actual emitter of the bands is the neutral BaO molecule.

The first attempt to make a vibrational quantum analysis of the bands of barium oxide was made by Mecke and Guillery[6] in 1927. They used, however, the older data, which are not in sufficiently close agreement with one another, as has been remarked by Kayser. Since then no further work has been done on these bands.

In the present investigation, the spectrum has been photographed under different dispersions and the measurements of band-heads extended in the region of the longer wave-lengths as far as λ 8000. Even with moderate dispersion the bands are partially resolved. In fact, their line structure was observed by several early investigators.

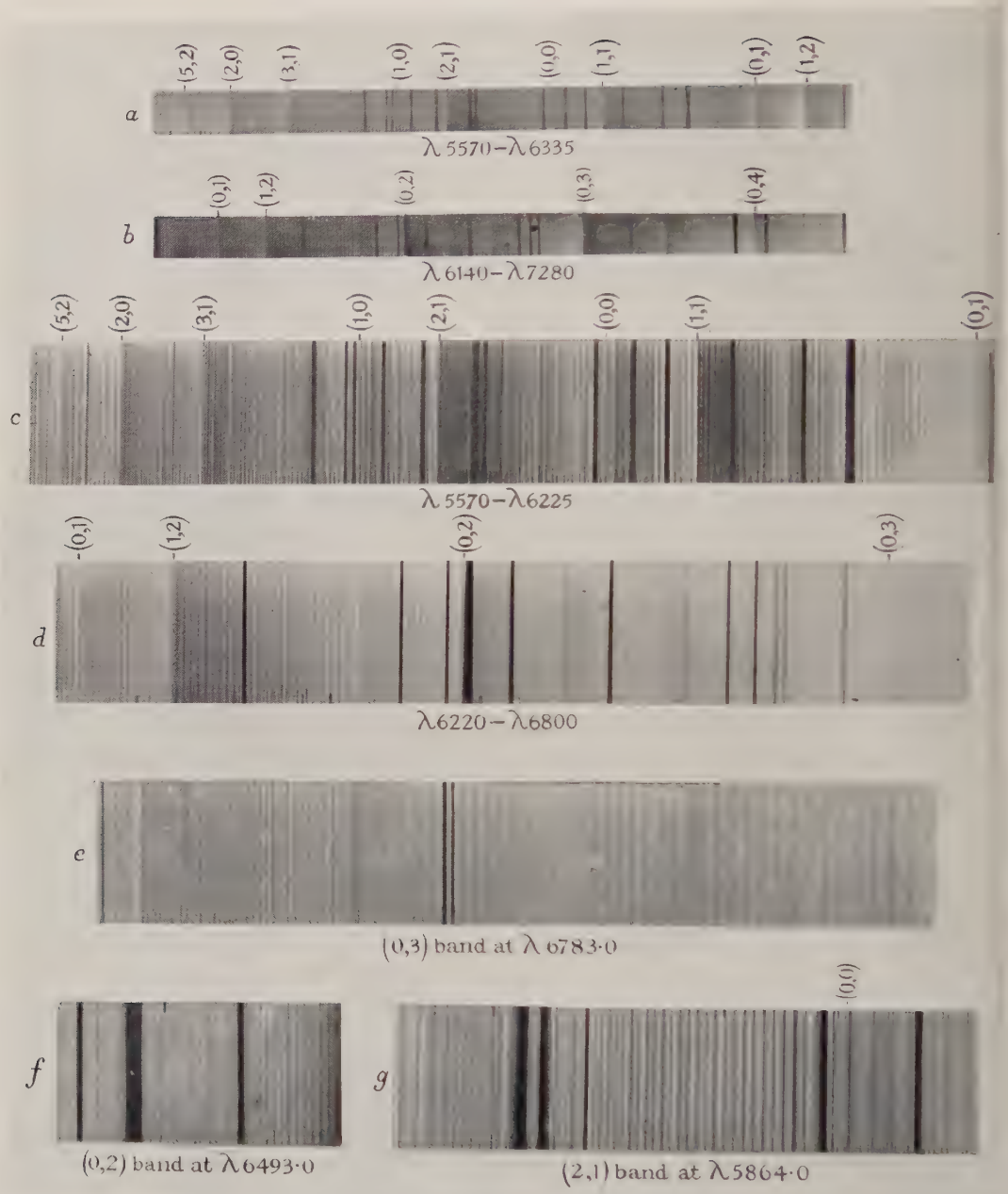
This paper has two objects. The first is to make a vibrational quantum analysis of the bands, based upon new measurements. It has been found that the bands between λ 8000–λ 4300 form a single system*. The second is to obtain the fine-structure analysis of this system.

§ 2. EXPERIMENTAL

In the present investigation the source employed is the flame surrounding a carbon arc with either $BaCl_2$ or $Ba(NO_3)_2$ in the lower (+) carbon, carrying a current of about eight amperes from 220-volt d.c. mains.

For a preliminary survey, spectrograms were taken with the Hilger Littrow-mounted E 1 quartz prism spectrograph in the region λ 7500 λ 2200. With Ilford Special Rapid Panchromatic plates, an exposure of about fifteen minutes was sufficient to get the bands strongly developed between λ 6800 -λ 4300. Longer exposures were necessary to extend them on both ends. The extension of the bands in the red region was therefore taken up with the help of a Hilger E 52 glass prism spectrograph by using plates sensitized with pinacyanol and with rubro-cyanine

* A preliminary note about this spectrum was published in *Nature*, **131**, 402 (1933), in which the bands were attributed to two systems, but there were a few fairly strong bands which could not be fitted into either of them. It was only on longer exposure that the nature of the rather faint band which forms the (0, 0) of the present system was disclosed, leading to the analysis into a single system which now includes all the observed bands. The intensity-distribution of the bands falls on a wide Condon parabola accounting for the weakness of the (0, 0) head. The result of the rotational analysis also supports this arrangement. The present paper, therefore, supersedes the above-mentioned note.



BaO bands; *a*, *b*, Hilger E 52 glass spectrograph; *c*, *d*, 6-ft. grating, first order; *e*, *f*, *g*, 15-ft grating, first order.

To face page 53

as well as Ilford Infra-red-sensitive and Kodak Kryptocyanine plates. Bands were present as far as λ 8400, which is the maximum limit of sensitivity that can be obtained ordinarily with these plates. The heads are single and fairly sharp except where overlapping is great. But the bands below λ 4300, besides being superposed upon the violet CN bands and the OH bands, were found to be very weakly developed.

For measurements of band-heads, besides the quartz and the glass prism spectrograms photographs of the bands in the region λ 6900–λ 4800 were taken in the first order of a 6-ft. Rowland concave grating set up in a Paschen mounting and having a dispersion of about 8·9 Å./mm. A part of the spectrum under different dispersions is reproduced in the plate.

For measurements of structure lines, spectrograms were taken in the first and second orders of a 15-ft. concave grating set up in a similar mounting. The dispersion is about 1·8 Å./mm. in the second order. There is a complete resolution except for a few lines near the heads of the bands. For each band analysed the lines were measured at least twice on different spectrograms. The measurements of individual lines are considered to be accurate to within 0·05 cm^{-1}.

A Gaertner comparator was used for measurements. For standards of comparison, Pfund iron arc lines and neon lines were employed. In some regions, barium arc lines were also used. Precise measurements of band-heads lying between λ 8400 and λ 8000, and those below λ 4300, could not be made owing to the low intensity.

The measurements on different spectrograms revealed that practically all the band-heads (with the exception of a few weaker ones) that have been measured by the early investigators between λ 7100–λ 4300 have come out in the present investigation. Some of the weaker bands, especially those of Börsch in the red region, were found to be mere condensations of structure lines presenting the appearance of a head. The data of Olmsted and of Börsch appeared, however, to be more in agreement with the present values than those of Eder and Valenta.

§ 3. VIBRATIONAL ANALYSIS

The wave-length and wave-number data of band-heads together with the assignment of their vibrational quantum numbers are given in table 1. The relative photographic intensity data of the bands have been taken from those given by Olmsted and by Börsch. In cases where new band-heads have been measured, their intensities have been visually estimated in comparison with those of others. These intensity data are enclosed in brackets.

The equation representing the wave-number of band-heads is

$$\nu = 16815{\cdot}89 + \{495{\cdot}36\,(v' + \tfrac{1}{2}) - 1{\cdot}12\,(v' + \tfrac{1}{2})^2\} - \{671{\cdot}48\,(v'' + \tfrac{1}{2}) - 2{\cdot}20\,(v'' + \tfrac{1}{2})^2\} \quad \ldots\ldots(1).$$

It may here be noted that the bands measured by the early investigators below λ 4300 probably belong to a different system.

Table 1. New data and vibrational quantum analysis

v'	v''	λ_{air}	$\nu_{vac.}$	v'	v''	λ_{air}	$\nu_{vac.}$
1	7	7905·10 (1)	12646·6	8	4	5562·39 (1)	973·9
0	6	7815·20 (3)	792·1	4	1	5547·82 (3)	18020·1
2	7	7610·10 (1)	13136·8	7	3	5509·73 (3)	144·7
1	6	7523·49 (2)	288·0	3	0	5492·69 (10)	201·0
0	5	7440·42 (4)	436·4	10	5	5470·50 (1)	274·8
3	7	7336·87 (1)	626·0	6	2	5454·61 (3)	328·0
2	6	7254·70 (2)	780·4	9	4	5419·77 (2)	445·9
1	5	7176·60 (3)	930·2	5	1	5402·68 (2)	504·2
0	4	7097·41 (5)	14085·8	8	3	5366·65 (2)	628·4
3	6	7007·15 (1)	267·2	4	0	5349·67 (8)	687·6
2	5	6931·43 (2)	423·1	7	2	5315·85 (1)	806·4
1	4	6857·22 (2)	579·2	6	1	5265·16 (1)	987·5
0	3	6782·79 (8)	739·1	9	3	5233·42 (1)	19102·7
3	5	6704·50 (2)	911·2	5	0	5214·68 (7)	171·3
2	4	6633·48 (2)	15070·9	8	2	5185·02 (1)	281·0
1	3	6563·20 (3)	232·3	7	1	5135·00 (1)	468·8
0	2	6493·09 (9)	396·7	10	3	5106·78 (1)	576·4
3	4	6423·12 (4)	564·5	6	0	5086·66 (6)	653·8
2	3	6358·05 (3)	723·8	9	2	5059·96 (1)	757·5
1	2	6291·01 (8)	891·3	8	1	5012·45 (2)	944·8
0	1	6224·66 (6)	16060·7	11	3	4988·06 (1)	20042·3
7	6	6169·58 (3)	204·1	7	0	4965·45 (3)	133·6
3	3	6165·11 (6)	215·8	10	2	4941·73 (2)	230·2
10	8	6116·20 (2)	345·2	9	1	4896·53 (2)	417·0
6	5	6111·00 (2)	359·4	12	3	4873·80 (4)	512·2
2	2	6102·32 (5)	382·7	8	0	4850·57 (6)	610·4
1	1	6039·64 (9)	552·7	11	2	4830·02 (2)	698·1
8	6	5996·60 (1)	671·5	10	1	4784·15 (4)	896·5
4	3	5984·86 (4)	704·2	13	3	4765·70 (2)	977·3
0	0	5976·28 (3)	728·1	9	0	4740·98 (4)	21086·8
7	5	5933·74 (1)	848·1	12	2	4722·73 (3)	168·3
3	2	5925·13 (2)	872·5	11	1	4680·27 (5)	360·3
10	7	5886·90 (4)	982·2	14	3	4663·50 (3)	437·1
6	4	5875·29 (2)	17015·7	10	0	4636·70 (4)	560·7
2	1	5864·51 (10)	047·0	13	2	4621·00 (4)	633·9
13	9	5839·98 (1)	118·6	12	1	4579·42 (4)	830·7
9	6	5829·20 (2)	150·3	11	0	4537·77 (2)	22030·0
5	3	5817·57 (2)	184·4	14	2	4524·01 (4)	098·1
1	0	5805·15 (6)	221·3	13	1	4483·80 (3)	296·3
8	5	5770·07 (3)	325·9	12	0	4442·61 (1)	503·0
4	2	5758·39 (3)	361·2	15	2	4430·60 (3)	564·0
7	4	5713·68 (3)	497·0	14	1	4391·26 (1)	766·1
3	1	5700·99 (8)	535·9	17	3	4381·20 (3)	818·4
10	6	5672·91 (4)	622·8	13	0	4353·90 (1)	961·5
6	3	5659·00 (2)	666·1	16	2	4343·30 (2)	23017·5
2	0	5644·13 (9)	712·7	15	1	4304·22 (1)	226·5
9	5	5615·42 (1)	803·1	14	0	4268·85 (2)	421·4
5	2	5602·37 (4)	844·7				

§ 4. ROTATIONAL ANALYSIS

The bands are of the simple two-branch type originating from a $^1\Sigma - {}^1\Sigma$ transition. Except for the (0, 2) band, the lines of the two branches form close doublets which coalesce into single lines as they approach the head; see (*e*) and (*g*) in the plate. In the (0, 2) band, apparently there is a single series of lines proceeding from the head. This is due to the superposition of the *R*-branch lines returning from

the head upon those of the same series proceeding towards the head from the origin as well as upon the lines of the P-branch as they run away from the origin; see (f) in the plate. Owing to high-temperature excitation of the bands, the series are long. No evidence was found of a Q branch, although a systematic search was made in every band for Q lines of low intensity. There is no sign of splitting into finer components even for last noticeable lines on the plates, and only a single line, $P(0)$, is missing.

By definition,

$$R(K) = T'(K+1) - T''(K) \quad \text{.....(2),}$$

$$P(K) = T'(K-1) - T''(K) \quad \text{.....(3),}$$

so that

$$\Delta_2 T'(K) = R(K) - P(K) \quad \text{.....(4),}$$

$$\Delta_2 T''(K) = R(K-1) - P(K+1) \quad \text{.....(5).}$$

According to the combination principle, all bands having the same upper vibrational state should yield sets of values of $\Delta_2 T'(K)$ which are numerically identical; and similarly all bands having the same lower vibrational state should give identical sets of values of $\Delta_2 T''(K)$.

The wave-numbers of the lines together with their quantum assignments for each of the bands analysed as well as the values of $\Delta_2 T'(K)$ and $\Delta_2 T''(K)$ are given in tables 2 to 6.

The next step was to calculate the rotational constants of the molecule from the combination differences. Since the rotational energy of a molecule in a $^1\Sigma$ state is $T(K)$, where

$$T(K) = B_v K(K+1) + D_v K^2 (K+1)^2 + \ldots,$$

the combination differences can be expressed thus:

$$\begin{aligned}\Delta_2 T(K) &= T(K+1) - T(K-1) \\ &= 4B_v (K+\tfrac{1}{2}) + 8D_v (K+\tfrac{1}{2})^3 + \ldots \quad \text{.....(6),}\end{aligned}$$

where terms small in comparison with $8D_v (K+\frac{1}{2})^3$ are dropped. Both B_v and D_v depend upon v according to the relations

$$\begin{aligned}B_v &= B_e - \alpha(v+\tfrac{1}{2}) \\ &= B_0 - \alpha v \quad \text{.....(7),}\end{aligned}$$

and

$$\begin{aligned}D_v &= D_e + \beta(v+\tfrac{1}{2}) \\ &= D_0 + \beta v \quad \text{.....(8).}\end{aligned}$$

Here B_e and D_e are the extrapolated values of B_v and D_v corresponding to the non-vibrating molecule.

The mean values of the combination differences $\Delta_2 T(K)$ have been taken in cases where there was more than one datum available for a particular pair of rotational levels. For each vibrational level, the values of $\Delta_2 T(K)$ were plotted against K. It was found that the points lay very approximately on a straight line, so that terms of higher power than the third need not be considered and the term involving the third power is itself very small. The slope of this line in each case gave also a good approximation to the value of $4B_v$.

Table 2. Structure of the (0, 3) band

K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ ($v'=0$)	$\Delta_2T''(K)$ ($v''=3$)	K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ ($v'=0$)	$\Delta_2T''(K)$ ($v''=3$)
0	—				51	36·38	73·14	63·24	73·51
1	—	14736·42	—	—	52	31·82	67·39	64·43	—
2	—	35·48	—	—	53	27·10	—	—	76·31
3	—	34·51	—	—	54	22·32	55·51	66·81	77·69
4	—	33·47	—	—	55	17·46	49·41	68·05	79·11
5	—	32·38	—	—	56	12·46	43·21	69·25	80·46
6	14739·13	31·10	8·03	—	57	07·42	37·00	70·42	81·87
7	39·13	29·82	9·31	10·79	58	02·22	30·59	71·63	83·27
8	38·90	28·34	10·56	12·22	59	596·95	24·15	72·80	84·69
9	38·68	26·91	11·77	13·68	60	91·54	17·53	74·02	86·10
10	38·26	25·22	13·04	15·09	61	86·05	10·85	75·20	87·46
11	37·85	23·59	14·26	16·53	62	80·47	04·07	76·40	88·86
12	37·24	21·73	15·51	17·97	63	74·78	497·19	77·59	90·28
13	36·61	19·88	16·73	19·39	64	14568·96	14490·19	78·77	91·64
14	35·76	17·85	17·91	20·80	65	63·10	83·14	79·96	93·01
15	34·92	15·81	19·11	22·20	66	57·10	75·95	81·15	94·40
16	33·90	13·56	20·34	23·64	67	51·02	68·70	82·32	95·79
17	32·86	11·28	21·58	25·11	68	44·86	61·31	83·55	97·18
18	31·60	08·79	22·81	26·51	69	38·53	53·84	84·69	98·64
19	14730·38	14706·35	24·03	27·95	70	32·13	46·22	85·91	99·97
20	28·93	03·65	25·28	29·35	71	25·65	38·56	87·09	101·34
21	27·54	01·03	26·51	30·77	72	19·09	30·79	88·30	102·69
22	25·91	698·16	27·75	32·20	73	12·40	22·96	89·44	104·09
23	24·31	95·33	28·98	33·67	74	05·60	15·00	90·60	105·48
24	22·46	92·24	30·22	35·09	75	498·70	06·92	91·78	106·86
25	20·64	89·22	31·42	36·50	76	91·70	398·74	92·96	108·23
26	18·60	85·96	32·64	37·94	77	84·61	90·47	94·14	109·59
27	16·57	82·70	33·87	39·39	78	77·39	82·11	95·28	111·06
28	14·32	79·21	35·11	40·85	79	70·04	73·55	96·49	112·37
29	12·06	75·72	36·34	42·22	80	62·65	65·02	97·63	113·65
30	09·64	72·10	37·54	43·62	81	55·16	56·30	98·77	115·12
31	07·19	68·44	38·75	45·05	82	47·54	47·53	100·01	116·52
32	04·58	64·59	39·99	46·47	83	39·81	38·64	101·17	117·84
33	01·96	60·72	41·24	47·91	84	32·00	29·70	102·30	119·20
34	699·15	56·67	42·48	49·41	85	24·11	20·61	103·50	120·59
35	96·28	52·55	43·73	50·83	86	16·10	11·41	104·69	122·00
36	93·26	48·32	44·94	52·29	87	07·99	02·11	105·88	123·29
37	14690·17	14643·99	46·18	53·71	88	399·80	292·81	106·99	124·64
38	86·95	39·55	47·40	55·12	89	91·51	83·35	108·16	—
39	83·67	35·05	48·62	56·55	90	83·12	—	—	—
40	80·26	30·40	49·86	57·98	91	14374·60	—	—	—
41	76·79	25·69	51·10	59·38	92	66·01	—	—	—
42	73·20	20·88	52·32	60·80	93	57·32	—	—	—
43	69·52	15·99	53·53	62·15	94	48·48	—	—	—
44	65·81	11·05	54·76	63·57	95	39·63	—	—	—
45	61·90	05·95	55·95	65·08	96	30·63	—	—	—
46	57·90	00·73	57·17	66·49	97	14321·55	—	—	—
47	53·80	595·41	58·39	67·87	98	12·31	—	—	—
48	49·63	90·03	59·60	69·33	99	03·01	—	—	—
49	45·27	84·47	60·80	70·75	100	293·63	—	—	—
50	40·90	78·88	62·02	72·13	101	84·11	—	—	—

These figures for B_v were then used to calculate approximately the values of D_v from several observed values of $\Delta_2T(K)$. Then by successive approximation and repeated trials values were finally assigned to B_v and D_v. The values of B_v thus found showed a linear variation according to equation (7). The values of D_v were

Table 3. Structure of the (0, 2) band

K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ $(v'=0)$	$\Delta_2T''(K)$ $(v''=2)$	K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ $(v'=0)$	$\Delta_2T''(K)$ $(v''=2)$
0	15395·10				37	45·65	299·45	46·20	53·89
1	95·65	15393·78	1·87	2·15	38	42·39	94·96	47·43	55·28
2	96·05	92·95	3·10	3·60	39	39·03	90·37	48·66	56·69
3	96·41	92·05	4·36	5·05	40	35·57	85·70	49·87	58·12
4	96·60	91·00	5·60	6·50	41	31·99	80·91	51·08	59·59
5	96·74	89·91	6·83	—	42	28·32	75·98	52·34	61·03
6	96·74	—	—	—	43	24·52	70·96	53·56	62·51
7	96·60	—	—	10·83	44	20·58	65·81	54·77	63·91
8	96·41	85·91	10·50	—	45	—	60·61	—	65·28
9	96·05	—	—	—	46	12·46	55·30	57·16	—
10	95·65	—	—	15·20	47	08·21	49·84	58·37	68·18
11	95·10	80·85	14·25	16·60	48	03·90	44·28	59·62	69·58
12	94·51	79·05	15·46	18·02	49	299·45	38·63	60·82	70·96
13	93·78	77·08	16·70	19·48	50	94·96	32·94	62·02	72·31
14	92·95	75·03	17·92	20·91	51	90·37	27·14	63·23	73·70
15	92·05	72·87	19·18	22·29	52	15285·70	15221·26	64·44	—
16	15391·00	15370·66	20·34	23·73	53	80·91	—	—	—
17	89·91	68·32	21·59	25·14	54	75·98	—	—	—
18	—	65·86	—	26·60	55	70·96	—	—	—
19	—	63·31	—	—	56	65·81	—	—	—
20	85·91	60·62	25·29	—	57	60·61	—	—	—
21	—	57·81	—	30·98	58	55·30	—	—	—
22	—	54·93	—	—	59	49·84	—	—	—
23	80·85	51·91	28·94	—	60	44·28	—	—	—
24	79·05	48·85	30·20	35·20	61	38·63	—	—	—
25	77·08	45·65	31·43	36·66	62	32·94	—	—	—
26	75·03	42·39	32·64	38·05	63	27·14	—	—	—
27	72·87	39·03	33·84	39·46	64	21·26	—	—	—
28	70·66	35·57	35·09	40·88	65	14·98	—	—	—
29	68·32	31·99	36·33	42·34	66	08·87	—	—	—
30	65·86	28·32	37·54	43·80	67	02·62	—	—	—
31	15363·31	15324·52	38·79	45·28	68	196·20	—	—	—
32	60·62	20·58	40·04	—	69	89·62	—	—	—
33	57·81	—	—	48·16	70	83·20	—	—	—
34	54·93	12·46	42·47	49·60	71	76·43	—	—	—
35	51·91	08·21	43·70	51·03	72	69·63	—	—	—
36	48·85	03·90	44·95	52·46					

found to be practically constant, so that β is negligibly small. Its value was, however, calculated from the theoretical formula

$$\beta = \frac{\alpha^2}{6\omega_e} + \frac{20\alpha B_e^2 - 32x_e B_e^3}{\omega_e^2} \qquad \text{.....(9).}$$

The following constants were thus evaluated:

$B_e' = 0{\cdot}30995$ cm.$^{-1}$
$B_0' = 0{\cdot}3094$
$B_1' = 0{\cdot}3083$
$B_2' = 0{\cdot}3072$
$\alpha' = 0{\cdot}0011$
$D_e' = -\,0{\cdot}483 \times 10^{-6}$
β' (calculated) $= 0{\cdot}24 \times 10^{-9}$
$r_e' = 1{\cdot}938 \times 10^{-8}$ cm.
$I_e' = 89{\cdot}44 \times 10^{-40}$ gm. cm.^{2}

$B_e'' = 0{\cdot}36440$ cm.$^{-1}$
$B_1'' = 0{\cdot}3620$
$B_2'' = 0{\cdot}3604$
$B_3'' = 0{\cdot}3588$
$\alpha'' = 0{\cdot}0016$
$D_e'' = -\,0{\cdot}430 \times 10^{-6}$
β'' (calculated) $= -\,1{\cdot}2 \times 10^{-9}$
$r_e'' = 1{\cdot}787 \times 10^{-8}$ cm.
$I_e'' = 76{\cdot}04 \times 10^{-40}$ gm. cm.^{2}

Table 4. Structure of the (1, 2) band

K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ (v' 1)	$\Delta_2T''(K)$ (v'' 2)	K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ (v' = 1)	$\Delta_2T''(K)$ (v'' = 2)
0	—				30	15859·40	15822·00	37·40	43·88
1	—	—	—	—	31	56·71	18·09	38·62	45·31
2	—	—	—	—	32	53·92	14·09	39·83	46·72
3	—	—	—	—	33	51·05	09·99	41·06	48·15
4	—	15886·04	—	—	34	48·06	05·77	42·29	49·61
5	—	84·70	—	—	35	45·00	01·44	43·56	51·01
6	15891·31	83·31	8·00	—	36	41·86	797·05	44·81	52·47
7	91·11	81·90	9·21	10·80	37	38·54	92·53	46·01	53·94
8	90·91	80·51	10·40	12·26	38	35·13	87·92	47·21	55·31
9	90·51	78·85	11·66	13·73	39	31·62	83·23	48·39	56·76
10	90·10	77·18	12·92	15·14	40	28·01	78·37	49·64	58·12
11	89·60	75·37	14·23	16·60	41	24·35	73·50	50·85	59·62
12	88·92	73·50	15·42	18·04	42	20·45	68·39	52·06	—
13	88·21	71·56	16·65	19·43	43	16·52	—	—	62·47
14	87·38	69·49	17·89	20·91	44	12·57	57·98	54·59	63·90
15	86·39	67·30	19·09	22·29	45	08·39	52·62	55·77	65·34
16	85·40	65·09	20·31	23·72	46	04·27	47·23	57·04	66·74
17	84·19	62·67	21·52	25·17	47	799·92	41·65	58·27	—
18	82·96	60·23	22·73	26·60	48	95·53	—	—	—
19	81·57	57·59	23·98	28·11	49	90·96	—	—	—
20	80·06	54·85	25·21	29·57	50	86·32	—	—	—
21	78·42	52·00	26·42	30·97	51	81·56	—	—	—
22	76·75	49·09	27·66	32·39	52	76·66	—	—	—
23	74·91	46·03	28·88	33·84	53	71·60	—	—	—
24	73·02	42·91	30·11	35·21	54	66·43	—	—	—
25	71·05	39·70	31·35	36·63	55	61·11	—	—	—
26	68·96	36·39	32·57	38·07	56	55·70	—	—	—
27	66·77	32·98	33·79	39·57	57	50·21	—	—	—
28	64·41	29·39	35·02	40·99	58	44·58	—	—	—
29	61·97	25·78	36·19	42·41					

The difference between the observed and calculated values of $\Delta_2T'(K)$ and $\Delta_2T''(K)$ from equation (6) is not of any sensible magnitude even for large quantum numbers. This indicates that the coupling conforms to Hund's case (b) as classified by Mulliken[7].

A check on the correctness of the analysis is afforded by the theoretical relation

$$\omega_e^2 = -4B_e^3/D_e \qquad \text{.....(10).}$$

On substitution of the proper values of B_e and D_e, it is found that $\omega_e' = 496$ cm.$^{-1}$ and $\omega_e'' = 671$ cm.$^{-1}$ These values are in sufficiently close agreement with those obtained from the vibrational analysis of the bands.

A further check is afforded by Birge's rule[8]:

$$2x_eB_e/\alpha = 1{\cdot}4 \pm 0{\cdot}2 \qquad \text{.....(11).}$$

For these bands substitution of the proper values shows that

$$2x_e'B_e'/\alpha' = 1{\cdot}27 \quad \text{and} \quad 2x_e''B_e''/\alpha'' = 1{\cdot}49 \qquad \text{.....(12).}$$

A third check is obtained from the approximate relation between ω_e and r_e as given by Morse[9]:

$$\omega_e r_e^3 = \text{constant} \qquad \text{.....(13).}$$

Table 5. Structure of the (1, 1) band

K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ ($v'=1$)	$\Delta_2T''(K)$ ($v''=1$)	K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ ($v'=1$)	$\Delta_2T''(K)$ ($v''=1$)
0	—				30	16519·42	16481·97	37·45	44·09
1	—	—	—	—	31	16·69	77·98	38·71	45·54
2	—	—	—	—	32	13·88	73·88	40·00	46·96
3	—	—	—	—	33	10·93	69·73	41·20	48·39
4	—	—	—	—	34	07·88	65·49	42·39	49·80
5	—	16546·08	—	—	35	04·72	61·13	43·59	51·30
6	16552·71	44·71	8·00	—	36	.01·42	56·58	44·84	52·80
7	52·51	43·31	9·20	10·85	37	497·99	51·92	46·07	54·21
8	52·30	41·86	10·44	12·29	38	94·54	47·21	47·33	55·62
9	51·91	40·22	11·69	13·75	39	90·89	42·37	48·52	57·12
10	51·48	38·55	12·93	15·22	40	—	37·42	—	58·50
11	50·87	36·69	14·18	16·73	41	83·41	32·39	51·02	—
12	50·16	34·75	15·41	18·12	42	79·50	27·26	52·24	61·41
13	49·39	32·75	16·64	19·57	43	75·48	22·00	53·48	62·94
14	48·46	30·59	17·87	20·99	44	71·27	16·56	54·71	64·40
15	47·47	28·40	19·07	22·35	45	66·93	11·08	55·84	65·73
16	46·41	26·11	20·30	23·79	46	62·59	05·54	57·05	—
17	45·23	23·68	21·55	25·27	47	58·26	—	—	—
18	43·91	21·14	22·77	26·74	48	53·71	—	—	—
19	42·49	18·49	24·00	28·17	49	48·94	—	—	—
20	40·96	15·74	25·22	29·65	50	44·07	—	—	—
21	39·29	12·84	26·45	31·12	51	39·18	—	—	—
22	37·51	09·84	27·67	32·55	52	34·18	—	—	—
23	35·65	06·74	28·91	33·99	53	29·05	—	—	—
24	33·65	03·52	30·13	35·45	54	23·70	—	—	—
25	31·55	00·20	31·35	36·91	55	18·25	—	—	—
26	29·31	496·74	32·57	38·35	56	12·70	—	—	—
27	27·01	93·20	33·81	—	57	06·94	—	—	—
28	24·61	—	—	41·15	58	01·10	—	—	—
29	22·07	85·86	36·21	42·64					

It has been well established that this constant has a value equal to 3000×10^{-24} cm.^{2} in the cases of diatomic molecules composed of atoms of nearly equal mass. But if the masses of the atoms are unequal, as in the present case, the value of this constant should be greater than 3000×10^{-24} cm.^{2} This condition is satisfied since we have

$$\omega_e' r_e'^3 = 3605 \times 10^{-24} \text{ cm.}^2 \text{ and } \omega_e'' r_e''^3 = 3830 \times 10^{-24} \text{ cm.}^2 \quad \text{.....(14).}$$

The heats of dissociation for the upper as well as for the lower states have been calculated from the expression

$$D = \omega_e^2 / 4x_e\omega_e \quad \text{.....(15),}$$

and are found to be 6·76 and 6·32 volts respectively.

From the relation

$$\nu_{e\,(\text{mol})} + D' = D'' + \nu_{(\text{atom})} \quad \text{.....(16),}$$

$\nu_{(\text{atom})}$ is calculated to be 2·52 volts.

It may here be pointed out that Mulliken[10] in discussing the constitution of BeO molecule has suggested its normal state to be derived from Be (^{1}S) + O (^{1}D) atoms. For the homologous molecule BaO, it may not be unjustifiable to suppose that if the lower state of the observed band system is its normal state it will dissociate into B (^{1}S) and O (^{1}D) atoms, and in the upper state it will dissociate into

Table 6. Structure of the (2, 1) band

K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ $(v'-2)$	$\Delta_2T''(K)$ $(v''-1)$	K	$R(K)$	$P(K)$	$\Delta_2T'(K)$ $(v'=2)$	$\Delta_2T''(K)$ $(v''=1)$
0	—				38	84·28	37·35	46·93	55·53
1	—	—	—	—	39	80·41	32·44	47·97	57·00
2	—	17043·49	—	—	40	76·58	27·28	49·30	—
3	—	42·51	—	—	41	—	—	—	59·84
4	—	41·40	—	—	42	68·47	16·71	51·76	—
5	17046·01	40·24	6·77	—	43	64·13	11·24	52·89	62·66
6	—	38·99	—	9·41	44	59·76	05·81	53·95	63·95
7	—	37·60	—	—	45	—	00·18	—	65·32
8	—	36·03	—	—	46	50·76	894·44	56·32	—
9	—	34·40	—	—	47	46·05	88·59	57·45	68·27
10	17045·32	32·43	12·89	—	48	41·28	82·49	58·79	69·75
11	44·71	30·63	14·08	16·69	49	36·40	76·30	60·10	71·13
12	43·93	28·63	15·30	18·11	50	31·35	70·15	61·20	72·56
13	43·00	26·60	16·40	19·56	51	26·22	63·84	62·38	74·00
14	41·92	24·37	17·55	20·98	52	20·98	57·35	63·63	75·43
15	40·79	22·02	18·77	22·42	53	15·62	50·79	64·83	76·89
16	39·51	19·50	20·01	23·87	54	10·19	44·09	66·10	78·23
17	38·15	16·92	21·23	25·28	55	04·64	37·39	67·25	79·69
18	36·70	14·23	22·47	26·73	56	898·98	30·50	68·48	81·13
19	35·19	11·42	23·77	28·18	57	93·19	23·51	69·68	82·57
20	33·59	08·52	25·07	29·65	58	16887·31	16816·41	70·90	84·05
21	31·79	05·54	26·29	31·14	59	81·24	09·14	72·10	85·46
22	29·88	02·45	27·43	32·55	60	75·16	01·85	73·31	86·89
23	27·85	16999·24	28·61	34·02	61	68·85	794·35	74·50	88·32
24	25·70	95·86	29·84	35·49	62	62·56	86·84	75·72	89·84
25	23·50	92·36	31·14	36·92	63	55·96	79·01	76·95	91·46
26	21·12	88·78	32·34	38·36	64	49·31	71·10	78·21	—
27	18·66	85·14	33·52	39·78	65	42·81	—	—	—
28	16·08	81·34	34·74	41·19	66	35·98	—	—	—
29	17013·47	16977·47	36·00	—	67	16829·12	—	—	—
30	10·69	—	—	44·04	68	22·07	—	—	—
31	07·77	69·43	38·34	45·49	69	15·08	—	—	—
32	04·76	65·20	39·56	46·97	70	07·90	—	—	—
33	01·58	60·80	40·78	—	71	00·64	—	—	—
34	16998·37	—	—	49·82	72	793·21	—	—	—
35	95·02	51·76	43·26	51·29	73	85·63	—	—	—
36	91·52	47·08	44·44	52·74	74	77·97	—	—	—
37	87·97	42·28	45·69	54·17	75	70·17	—	—	—

Ba ($^1P^0$) and O (^1D) atoms. Such an assumption finds a support from the fact that from the data of energy states published by Bacher and Goudsmit[11], the difference of energy values between a Ba atom in $(6s^2)\,^1S$ state and that in a $(6s6p)\,^1P^0$ excited state is 2·23 volts, differing by about 0·3 volt from that calculated from the present data. Such a discrepancy may well be expected in a calculation where the dissociation energies have been derived by long extrapolation.

§ 5. ACKNOWLEDGMENTS

My best thanks are due to Prof. P. N. Ghosh for his helpful criticism and encouragement during the course of the investigation. I further beg to acknowledge my indebtedness to Prof. M. N. Saha, F.R.S., of the Allahabad University, for the loan of the 15-ft. grating without which the investigation could not have been carried out.

REFERENCES

(1) C. M. Olmsted. *Z. f. wiss. Photogr.* **4**, 255 (1906).
(2) L. Börsch. *Z. f. wiss. Photogr.* **7**, 297 (1909).
(3) W. N. Hartley. *Proc. R.S.* **79**, 242 (1907).
(4) J. M. Eder and E. Valenta. *Atlas Typischer Spektren*, Wien (1924).
(5) A. Harnack. *Phys. Z.* **15**, 578 (1924).
(6) R. Mecke and M. Guillery. *Phys. Z.* **28**, 514 (1927).
(7) R. S. Mulliken. *Rev. Mod. Phys.* **2**, 115 (1930).
(8) R. T. Birge. *Phys. Rev.* **31**, 919 (1927).
(9) P. M. Morse. *Phys. Rev.* **34**, 57 (1929).
(10) R. S. Mulliken. *Rev. Mod. Phys.* **4**, 67 (1932).
(11) R. F. Bacher and S. Goudsmit. *Atomic Energy States* (1932).

621.396.674

THE ACTION OF A TUNED RECTANGULAR FRAME AERIAL WHEN TRANSMITTING SHORT WAVES

By Professor L. S. PALMER, M.Sc., Ph.D.,

AND

D. TAYLOR, M.Sc., The University College, Hull.

Received March 3, 1933, *and in revised form August* 8, 1933. *Read October* 20, 1933.

ABSTRACT. The work described in this communication is a continuation of a paper by one of us (L. S. P.) on the action of a tuned rectangular frame aerial when receiving short waves. The present development deals with the somewhat similar problem of the action of a tuned rectangular frame aerial when transmitting short waves. Experiments are described which show that a variation of the frame-dimensions affects both the frame current and the radiation, and that these quantities will only attain their maximum values if the frame-dimensions are suitably adjusted.

A theory is developed to account for the observed phenomena, and equations are obtained the solutions of which give approximately the necessary conditions for maximum frame current and maximum radiation.

§ 1. INTRODUCTION

It has been known for some time that the action of a frame aerial when used with short waves comparable in length to the frame-dimensions is quite different from its action when used with long waves. In a recent investigation* one of us (L. S. P.) has shown that, when a tuned frame aerial comparable in dimensions to the wave-length is used for reception, the received current is very critically dependent on the frame-dimensions, and that by suitably varying the dimensions the current may be adjusted to a maximum. These facts were explained by taking into consideration the mutual action between the different limbs of the frame. Equations were developed which expressed the conditions for maximum frame current. These were eventually simplified to

$$\tan [a (1 + \cos \gamma) - \psi] = (a^2 - 1)/a \qquad \ldots\ldots(1),$$

$$\tan [a' (1 + \sin \gamma) + \psi] = (a'^2 - 1)/a' \qquad \ldots\ldots(2),$$

a, W, λ, a', H, γ, ψ

where $a = 2\pi W/\lambda$, $a' = 2\pi H/\lambda$, γ is the angle of incidence of the wave on the frame, and ψ is an arbitrary phase angle.

For convenience, this process of obtaining maximum frame current by the fulfilment of the above conditions has tentatively been called "formatization"† to distinguish it from the process of tuning, and the purpose of the present paper is to

* *Proc. R.S.* A 136, 193–209 (1932).

† *Loc. cit.* p. 208.

develop a similar theory of formatization of tuned rectangular frame aerials when used for transmission.

In a previous paper the electro-magnetic wave incident on the frame was considered to cause a frame current which can be resolved into an initial component due to the primary action of the wave, and a supplemental component due to the field of the current in the adjacent parts of the frame. It was then assumed that the condition for maximum frame current is that the initial and supplemental components shall be in phase*. It was thought that this assumption was justified provided that (i) the component currents are uniformly distributed and (ii) the amplitudes of the component currents are sensibly independent of the frame-dimensions.

The question of the resultant current-distribution has now been investigated† and it is found that the current-variation round a short-wave frame aerial depends on the ratio of the perimeter of the frame to the wave-length. Experiments show that the consequent lack of uniformity in the frame current does not vitiate the assumption that the frame current is a maximum when the supplemental and initial components are in phase. If we consider any infinitely small length of the wire, then the currents flowing in any such elemental length may be taken to be uniformly distributed throughout the length at any given instant, and the co-phasing of the supplemental and initial elemental currents leads to the same formatizing conditions as are given in equations (1) and (2) above, because the phase angles are independent of the current-amplitudes. It would therefore appear that the uniformity of the component currents throughout the whole length of any limb of the frame is not a necessary condition for the establishment of maximum frame current. The fact that the experimental points in figures 7, 8 and 13 lie close to the theoretical curves gives additional support to this conclusion.

The second assumption has been shown to be justified as long as the distances between adjacent limbs are not less than about $0{\cdot}2\lambda$‡. With these assumptions, equations (1) and (2) were developed; they give the critical values of H/λ and W/λ for maximum current in a tuned rectangular frame receiving from a given source in a given direction determined by the value of γ. Conversely, from the reciprocity principle, a frame will transmit most effectively to the same source along the same path but in the reverse direction, when it is similarly proportioned and when the same currents circulate in it.

If, when the frame is used for transmission, the currents are to be produced in it not by a passing wave but by a local oscillator, then the optimum frame-dimensions for maximum current may not be the same as when the necessary energy is absorbed from a passing wave. Consequently, when the frame current produced by the local oscillator is a maximum, the frame-dimensions may not be correct for optimum radiation in the required direction; and vice versa, when the frame

* Palmer and Honeyball, *Proc. Inst. Rad. Eng.* **20**, 1349 (1932).

† See pp. 76–84 of this volume.

‡ Palmer and Honeyball, *J. Inst. E.E.* **67**, 1046 (1929), and Ratcliffe, Vedy and Wilkins, *ibid.* 70 522 (1932).

is correctly proportioned for maximum radiation, the current in it may not be a maximum.

In view of these considerations, it seems desirable to study first the variation, as the frame-dimensions are changed, of the frame current produced by a local oscillator and then to proceed to the more complicated conditions for maximum radiation.

Therefore, in the next section the conditions necessary for the production of maximum current in the transmitting frame will be considered, and afterwards the more complicated conditions for maximum radiation will be studied.

§ 2. THEORETICAL CONDITIONS FOR MAXIMUM FRAME CURRENT

Consider a rectangular frame directly coupled to, and actuated by, a local valve oscillator. The applied e.m.f., assumed constant, will generate a local current which will eventually affect all parts of the frame because they are influenced by the electromagnetic field set up, and also because they are metallically interconnected to form a closed circuit. Hence the current in any limb of the frame is the resultant of three components, namely (i) that produced directly by the transmitter, which we may call the *initial* current; (ii) that produced indirectly by waves travelling across the space between parallel adjacent wires, which, for convenience, we may call the *first supplemental* current; and (iii) that produced indirectly by waves travelling round the perimeter of the frame, which we may call the *second supplemental* current. Only the first two were considered in the previous papers.

Since the magnitude and the phase of the supplemental components are dependent on the frame-dimensions, it follows that the resultant current is also dependent on the frame-dimensions, and that by suitable variation of the dimensions the frame current can be adjusted to a maximum. In this section, the conditions necessary for the establishment of this maximum frame current, when the frame has already been tuned in the ordinary way, will be considered.

Since the amplitudes of the component currents are sensibly independent of the dimensions of the frame as long as these exceed $0{\cdot}2\lambda$, the condition for maximum frame current is that the current components shall be in phase. In order to develop the necessary equations, the phase of the supplemental components must be calculated.

First formatizing conditions. Consider the first supplemental current component due to the presence of the initial current in, say, an element of one of the horizontal limbs of the frame. Let the initial current be $I_0 \sin \omega t$ and let it be uniform throughout the element.

I_0, ω, t

Then the electric moment is given by

$$-\int I_0 \sin \omega t \, dt = (I_0 \cos \omega t)/\omega.$$

This results in a field a distance H away such that the horizontal electric force is given by

E, A, v

$$E = -AI_0 \left\{ \frac{1}{H^3} \cos(\omega t - a') + \frac{1}{vH^2} \frac{d}{dt} [\cos(\omega t - a')] + \frac{1}{v^2 H} \frac{d^2}{dt^2} [\cos(\omega t - a')] \right\}$$

$$= \frac{AI_0}{H^3} \sqrt{\{(a'^2 - 1)^2 + a'^2\}} \cdot \sin(\omega t - a' + \phi'),$$

or $$E_0 \sin(\omega t - a' + \phi'),$$

where $$E_0 = \frac{AI_0}{H^3}\sqrt{\{(a'^2 - 1)^2 + a'^2\}} \quad \text{and} \quad \tan\phi' = (a'^2 - 1)/a'. \qquad E_0, \phi'$$

If the frame be tuned, the current produced by this e.m.f. will be in phase with it, and therefore the first supplemental current in one horizontal element will differ in phase from the initial current in the opposite element by

$$-a' + \phi'.$$

Hence, for the first supplemental current to be in phase with the initial current in the same element, the energy must be re-radiated from the opposite element. Consequently reinforcement will occur if

$$2(-a' + \phi') = 0 \text{ or } 2n\pi \qquad \ldots\ldots(3),$$

where n is any integer. n

A precisely similar argument shows that the condition for these two current components to be in phase in the vertical members of the frame is

$$2(-a + \phi) = 0 \text{ or } 2n\pi \qquad \ldots\ldots(4),$$

where $$\tan\phi = (a^2 - 1)/a. \qquad \phi$$

These equations reduce to

$$(a^2 - 1)/a = \tan(n\pi + a) = \tan a \qquad \ldots\ldots(5)$$

and $$(a'^2 - 1)/a' = \tan(n\pi + a') = \tan a' \qquad \ldots\ldots(6).$$

These transcendental equations have been solved by a graphical method and the solutions are

$$a = a' = 4{\cdot}45,\ 7{\cdot}70, \text{ etc.},$$

i.e. $$W/\lambda = H/\lambda = 0{\cdot}71,\ 1{\cdot}20, \text{ etc.}$$

There is an ambiguity in these solutions, because equations (5) and (6) also represent the conditions for minimum frame current. Experiments show, however, that alternate solutions give maximum and minimum values respectively.

Thus we conclude that a frame of height $0{\cdot}71\lambda$ and width $0{\cdot}71\lambda$ will have an abnormally large current circulating round it when directly coupled to an oscillator. This result may be compared with that for maximum receiving current, namely, that when the frame-height was $0{\cdot}71\lambda$ the optimum width was either $0{\cdot}33\lambda$ or $0{\cdot}85\lambda$, and the frame was not square as in the present case.

Now the early experiments showed that if the critical width be altered, then a large current could still be maintained by suitably varying the height. This can be explained, as in the previous paper*, by extending the above reasoning as follows:

Suppose the frame-width be increased so that the phase angle between the component currents is not $2n\pi$ but $(2n\pi - 2\psi)$, say, then

$$2(-a + \phi) = 2n\pi - 2\psi \qquad \ldots\ldots(7).$$

The new width may be considered to act as an added inductance producing the undesirable phase-lag ψ. If now the height be decreased just sufficiently to produce a phase lead of $(2n\pi + 2\psi)$ in the currents in the horizontal wires of the frame, then

* *Proc. R.S.* A **136**, 196 (1932).

the effect on the frame current is similar to that of introducing a capacity of sufficient magnitude to compensate for the inductive effect of increasing the frame-width. Thus

$$2(-a' + \phi') = 2n\pi + 2\psi \qquad \text{.....(8).}$$

Equations (7) and (8) reduce to

$$(a^2 - 1)/a = \tan(n\pi + a - \psi) = \tan(a - \psi) \qquad \text{.....(9)}$$

and

$$(a'^2 - 1)/a' = \tan(n\pi + a' + \psi) = \tan(a' + \psi) \qquad \text{.....(10).}$$

The frame-dimensions (that is the value of H/λ and W/λ) for different values of ψ can be calculated from the above transcendental equations. This has been done by a graphical method and the relation between H/λ and W/λ is shown by the curve marked (*a*) in figure 1.

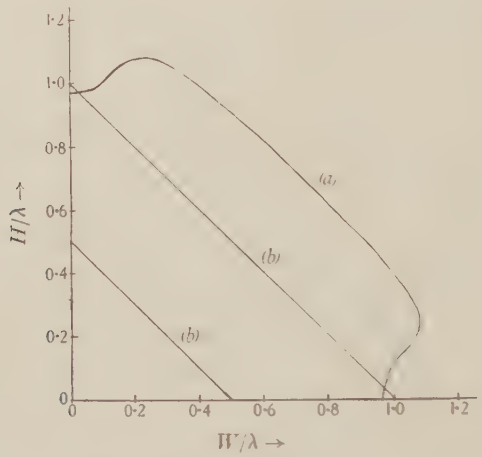


Figure 1.

Thus any point on this graph gives the frame-dimensions in terms of H/λ and W/λ for which the resultant current will be a maximum. For convenience these conditions will be called the *first formatizing conditions**.

In general, they will be quite different from those conditions necessary to make the second supplemental current be in phase with the initial current. Consequently, there will be a further set of conditions (conveniently referred to as the *second formatizing conditions*) depending on the critical dimension for the co-phasing of these two components. These conditions will now be considered.

Second formatizing conditions. In any element of the frame, let the initial current be $I_0 \sin \omega t$ (say): the second supplemental current will be that due to the return of the wave to the element under consideration after it has travelled round the frame. With a frame of negligible resistance the reduction in amplitude, as the wave travels round the frame, will be due mainly to the loss of energy by

* The same conditions may be deduced by considering the effect of an infinite number of re-radiations to and fro between opposite limbs of the frame.

radiation. It is shown below in §§ 4, 5 that for maximum current under these conditions the radiation is zero, and hence the resultant current will be given by

$$I_0 \sin \omega t + I_0 \sin (\omega t - 2\pi l/\lambda),$$

where l is the frame-perimeter and therefore equal to $2(H + W)$. The amplitude of the resultant current is $2I_0 \cos(\pi l/\lambda)$ which is a maximum when

$$\pi l/\lambda = 0 \quad \text{or} \quad n\pi,$$

where n is an integer. Hence, since $a = 2\pi W/\lambda$ and $a' = 2\pi H/\lambda$, the condition for maximum current reduces to

$$a + a' = n\pi \qquad \ldots\ldots(11)^*.$$

If, as before, we take W/λ to be abscissae and H/λ to be ordinates, then equation (11) is represented by a series of straight lines cutting the axes at 45°. These are the curves (*b*) in figure 1. Hence we must conclude that if the frame-dimensions satisfy a point on any of the curves drawn in figure 1, then an abnormally large current will circulate round the frame.

Equation (11) may be written

$$l = n\lambda.$$

Thus maximum frame current occurs when the frame-perimeter is an integral number of wave-lengths. This is also the condition for the formation of stationary waves such that the current nodes and antinodes are equally spaced round the frame, the number of nodes or antinodes being $2n$.

The experiments which were carried out to test the above theory will now be described.

§ 3. FRAME CURRENT MEASUREMENTS

Apparatus. Since the foregoing theory applies to frames comparable in dimensions with the wave-length, it is necessary to work with short waves, so that the size of the frame can be within reasonable limits. The oscillator had a wave-length range of 5·5 to 11·4 metres. The circuit is shown in figure 2, whilst figure 3 shows

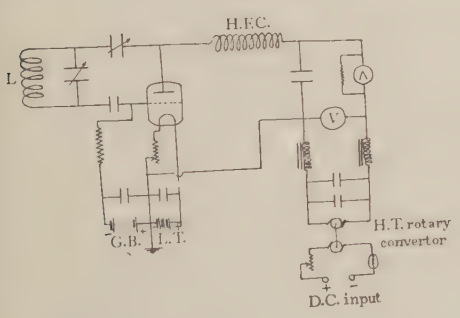


Figure 2.

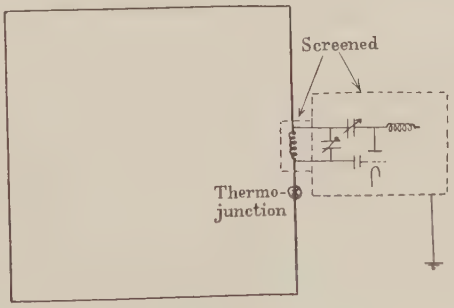


Figure 3.

the method by which the inductance L was directly coupled to the frame aerial. This inductance was a short strip of copper and formed a part of the frame. Because

* The same condition may be deduced by considering the effect produced by an infinite number of waves travelling round and round the frame. This method is employed in the following paper, pp. 76–84 of this issue.

of the small reactance-value of this strip compared with that of the frame, the tuning of the frame was not appreciably altered when the size of the frame was changed. The whole of the oscillator, including the inductance L, was screened by enclosing it in a sealed metal box. The current was measured with a thermo-junction and microammeter, which was read from a distance through a telescope.

The frame could be expanded or contracted in either or both dimensions by a modification of the method described by Palmer and Honeyball*. The new mechanism was so devised that all the necessary changes could be carried out from a distance of several wave-lengths. Errors arising from the movement of people in the vicinity of the aerial were thus obviated.

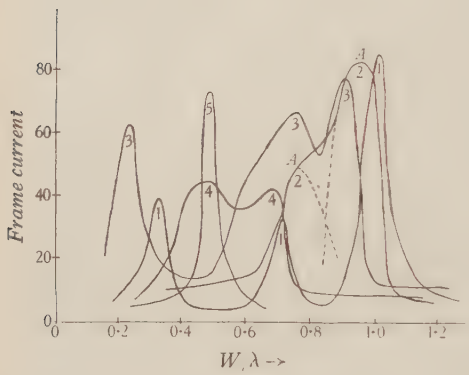


Figure 4. $\lambda=5{\cdot}85$ m. (1) $H/\lambda=0{\cdot}14$; (2) $H/\lambda=0{\cdot}35$; (3) $H/\lambda=0{\cdot}52$; (4) $H/\lambda=0{\cdot}72$; (5) $H/\lambda=0{\cdot}94$.

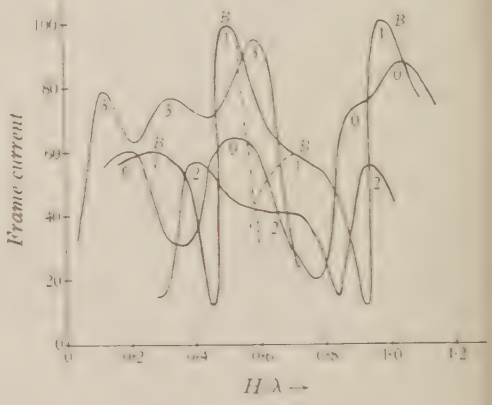


Figure 5. $\lambda=5{\cdot}9$ m. (0) $W/\lambda=0{\cdot}17$; (1) $W/\lambda=0{\cdot}36$; (2) $W/\lambda=0{\cdot}51$; (3) $W/\lambda=0{\cdot}83$.

Experimental measurements. In order to test the theory outlined in § 2 three methods were adopted. The general procedure in the first method was to take readings of the frame current for various widths, keeping the height of the frame and the radiated wave-length constant. Usually, readings were observed for every 20-cm. variation of the width. The measurements were repeated for frames of different heights varying from 1 to 6 metres. Figure 4 shows a set of results obtained in this way.

The procedure in the second method was exactly the same as in the first, but instead of the width, the height was varied. Some results obtained by this method are depicted in figure 5.

In the third method the frame current was recorded as before, but the size of the frame was kept fixed whilst the length of the radiated wave was varied by altering the tuning condenser of the oscillator. The wave-length was determined by an absorption wave-meter which had been calibrated from Lecher wires. Figure 6 shows the kind of results obtained, but these curves are less reliable than those shown in figures 4 and 5 because it was necessary for the operator to work near the aerial.

* *Proc. Inst. Radio Eng.* **20**, 1357 (1932).

The curves in figures 4, 5 and 6 illustrate the very marked current changes obtained. Examination of figures 4 and 5 shows that the curves obtained in this way are often far from regular. The curve marked *A*, figure 4, obtained with a height equal to $0{\cdot}35\lambda$, and the curve marked *B*, figure 5, obtained with a width equal to $0{\cdot}36\lambda$, are good examples of this irregularity. The irregularity can be explained by resolving such peaks into two smooth curves, and it is then found that one peak corresponds to the position given by the first formatizing conditions and the other peak corresponds to the position given by the second formatizing conditions. Thus the positions of the resolved peaks appear to conform to the theoretical requirements for maximum current. Hence in order to test the theory given in § 2 it will be necessary to resolve all such irregular curves into their

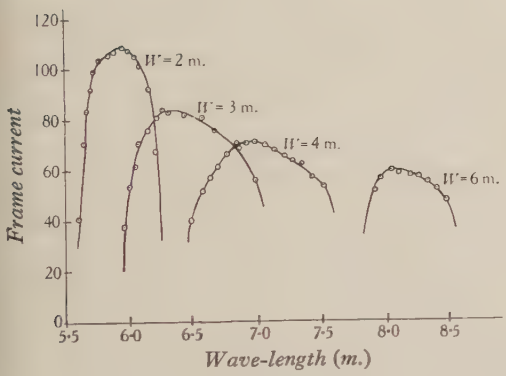


Figure 6. Height of frame, 5 m.

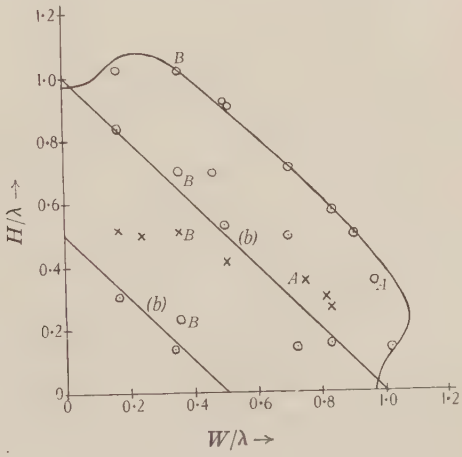


Figure 7.

components. The actual components in these two cases are shown dotted. This particular peculiarity is interesting in that it enables one to say which is the more potent factor in producing a large current—the first formatizing condition or the second formatizing condition. In every case it will be found that the first formatizing condition is by far the more important in these experiments.

It remains to be seen whether the actual positions of all the peaks approximate to those predicted by the theory. To test this point the peak values of H/λ and W/λ were calculated and plotted. The requisite data were obtained from figures 4, 5 and 6. Figure 7 shows the result of this analysis. For comparison, the theoretical curves of figure 1 have been added to the diagram. From the positions of the points marked with circles it appears that the theory given in § 2 is approximately correct. Experimental points have been found which lie on all the theoretical curves given in figure 1. In addition, however, certain experimental points, marked ×, do not lie on any of the theoretical curves. The explanation of these points appears to be as follows.

Reflection effects. In the theoretical discussion in § 2 the frame was considered

to be well removed from the influence of all other bodies. This does not obtain in the present instance, because the bottom of the frame was only about four feet above the ground. Since the ground, especially when wet, is a conductor, it will reflect electromagnetic waves incident upon it. Hence it is possible that the waves radiated from the frame may be reflected at the earth's surface and then received again by the frame. Under these circumstances the waves reflected back to the frame will be incident at an angle γ equal to 90° (with the nomenclature of the previous paper*), and consequently the frame will receive these reflected waves best (that is to say it will be formatized for these waves) when the frame-dimensions satisfy equations (1) and (2) in which $\gamma = 90^\circ$. With this condition these two equations reduce to

$$\tan(a - \psi) = (a^2 - 1)/a \qquad \text{.....(12)}$$

and

$$\tan(2a' + \psi) = (a'^2 - 1)/a' \qquad \text{.....(13).}$$

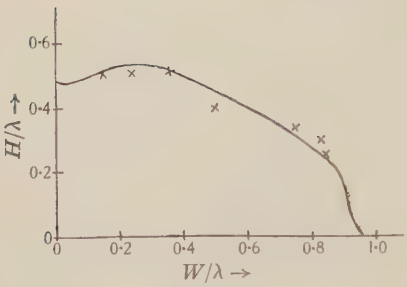


Figure 8.

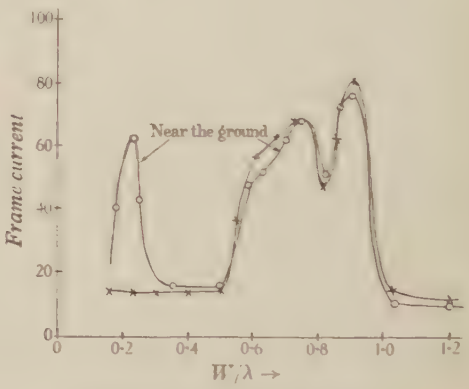


Figure 9. $\lambda = 5{\cdot}85$ m. $H = 0{\cdot}51\,\lambda$.

The solutions of these equations for different values of ψ are plotted in figure 8. In the same figure are shown the unexplained experimental points of figure 7, and it seems apparent from the graph that these experimental points are due to the presence of waves reflected off the ground. In order to test the point fully, however, some additional experiments were carried out.

The procedure in these experiments was to record the variation of frame current with changing width as the frame of constant height was elevated to different distances from the ground. The experimental arrangements allowed the whole frame and oscillator to be raised to various heights up to 16 m. In this way, true current-maxima due to the formatizing conditions would remain unaltered, whilst effects due to reflection from the ground should become less and less marked as the frame is elevated.

The results of such an experiment for a frame of height $0{\cdot}51\,\lambda$ are shown in figure 9. From these graphs it is apparent that the elevation was sufficient to cause one peak to disappear completely; the peak in question is marked by the second cross from the left in figures 7 and 8. In other experiments these subsidiary peaks

* *Loc. cit.*

could be intensified by artificially soaking the ground with water. The wetter the surface the stronger was the peak, whilst the other peaks remained unaffected. This is in agreement with the suggestion that these subsidiary peaks are caused by waves reflected from the ground.

Results. The results of the three experimental methods described in this section may now be summarized.

Experiments involving the variation of either the height or width of the frame or the wave-length lead to the same conclusion; namely, that maximum frame current is produced when either the first or second formatizing conditions are satisfied; i.e. when either equations (9) and (10) are satisfied or equation (11) is satisfied. In addition, the reflection experiments show that the frame current may also be a maximum when the conditions given by equations (12) and (13) are fulfilled.

§ 4. THEORETICAL CONDITIONS FOR MAXIMUM RADIATION

In the present paper consideration will be restricted to the case of transmission along the earth's surface in the direction of the plane of the frame. With this limitation the conditions necessary for maximum radiation from a transmitting frame are (1) large frame current, and (2) the correct co-phasing of the component fields at the particular point P (say) on the earth's surface at which the radiation is being received.

Conversely, the radiation to P will be small if either the frame current be small or the fields at P due to the currents in the several limbs of the frame be not in phase.

Equations (9) and (10) and equation (11) give the conditions for maximum frame current, whilst equations (1) and (2), with γ equal to 0 since P is on the earth's surface, give by the reciprocity theorem the conditions for the component fields at P to be in phase. The graphs of all these equations are shown in figure 10, which may be compared with figure 1. From figure 10 it is apparent that equations (9) and (10) and equations (1) and (2) are simultaneously satisfied only at the points of intersection A. Consequently, only for a transmitting frame of these particular dimensions (namely one in which $H/\lambda = 0{\cdot}40$ and $W/\lambda = 1{\cdot}00$ for the upper point and $H/\lambda = 0{\cdot}09$* and $W/\lambda = 1{\cdot}00$ for the lower point) will the radiation be good. For other frame-dimensions given by equations (9) and (10) the current will be large but the fields due to the current in parallel limbs will not completely reinforce at P, and hence the received radiation at P will be less.

Turning now to equation (11) which gives the condition for stationary waves; the graph of this equation has also solutions in common with those of equations (1) and (2), but in this case radiation is not to be expected as a result of the fulfilment of the stipulated conditions. The reason for this has been discussed by Macdonald†

* This dimension is less than 0·2 and consequently is outside the limits of accuracy of the present theory.

† Macdonald, *Electric Waves*, p. 62 *et seq.* (Cambridge University Press).

who shows that, when the frame-perimeter is an integral number of wave-lengths, then the fundamental free period of the frame is the same as the period of the applied e.m.f. The resulting oscillations are said to be *permanent* and there is no radiation from the frame. This will be the case whether the conditions given by equations (1) and (2) are fulfilled or not, and hence no radiation will be apparent when such permanent oscillations are present in the frame.

Owing to the reflection effects discussed in § 3, the transmitting frame will also have a large current in it when its dimensions satisfy equations (1) and (2) in which $\gamma = 90°$, i.e. when equations (12) and (13) are satisfied. If the graphs of these equations be plotted along with that for maximum radiation along the ground—i.e.

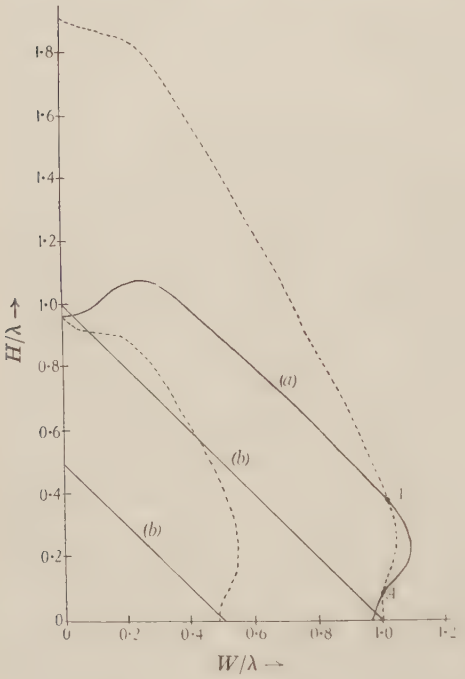


Figure 10.

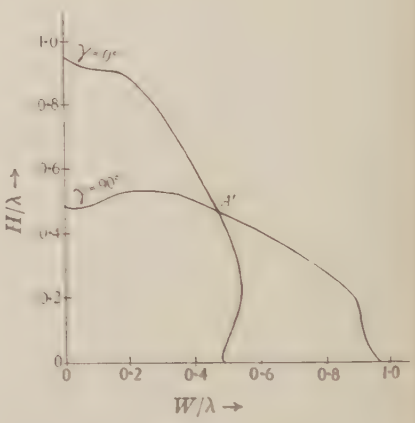


Figure 11.

equations (1) and (2) when $\gamma = 0$ —then it will be seen that at the point A' in figure 11 the two conditions are satisfied simultaneously. Thus, for the square frame of these particular dimensions (namely, $H = W = 0{\cdot}47\lambda$) a large radiation field may be expected at points along the earth's surface if reflected waves are appreciably affecting the frame current.

§ 5. HERTZIAN-RECEIVING-ROD MEASUREMENTS

The conditions for maximum radiation were tested by measuring the radiation from the frame with a Hertzian-rod receiver. The receiver used was similar to that already described by one of us*, and was situated in the same vertical plane as

* Palmer and Honeyball, *J. Inst. E.E.* **67**, 1049 (1929).

the frame at a distance of several wave-lengths. The microammeter deflections were read by means of a telescope by an observer lying on the ground at the foot of the receiver.

For constant frame-widths, the transmitting-frame current and the current received by the tuned Hertzian rod were recorded for different values of the frame-height, the wave-length remaining constant. Each set of readings were repeated for frame-widths of from 1 to 6 metres. The frame-current variations have already been discussed in the previous sections, and figure 12 shows a typical set of Hertzian-rod current-values plotted against the varying height of the transmitting frame. The critical frame-dimensions, in terms of H/λ and W/λ, for which the Hertzian-rod currents were a maximum, are plotted in figure 13 together with the theoretical curves from figures 10 and 11. The points in figure 13 are numbered to correspond with the peaks in figure 12 to which they refer.

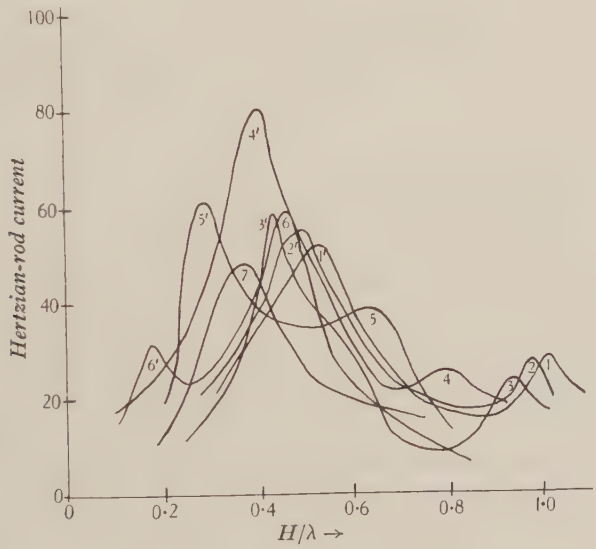


Figure 12. Radiation-measurements.

$\lambda = 5{\cdot}9$ m. (1) $W/\lambda = 0{\cdot}17$; (2) $W/\lambda = 0{\cdot}36$; (3) $W/\lambda = 0{\cdot}497$; (4) $W/\lambda = 0{\cdot}52$; (5) $W/\lambda = 0{\cdot}66$; (6) $W/\lambda = 0{\cdot}88$; (7) $W/\lambda = 1{\cdot}04$.

The conclusions arising from a consideration of these two figures may now be summarized as follows. (1) There are in general, for each value of the frame-width, two critical values of the height less than one wave-length for which the radiation is a maximum. The taller frame has the critical dimensions predicted by the first formatizing conditions, whilst the dimensions of the shorter frame are such that it is formatized for the reception of waves reflected from the ground. (2) There are no experimental points on the curves (*b*). That is to say, there is no radiation from frames the dimensions of which fulfil the second formatizing condition, or the perimeters of which are integral numbers of wave-lengths. If the Hertzian-rod currents are plotted on the same graph as the transmitting-frame currents, as in

figure 14, the absence of radiation when such stationary waves occur in the frame is more readily seen than from such curves as those of figure 12 where the Hertzian-rod currents alone are depicted. (3) Peaks 6 and 4′ in figure 12 show that the greatest radiation occurs from the frames the dimensions of which are given by the co-ordinates of these particular points in figure 13. The greatest radiation was anticipated on pages 71 and 72 from frames the dimensions of which are given by points A and A'; and it is significant that points 6 and 4′ lie close to these respective theoretical positions. Actually points 7 and 3′ lie somewhat closer to A and A'*, but it is thought that this discrepancy is an experimental error arising from the

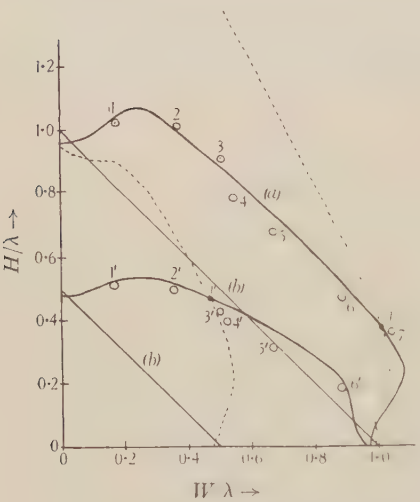


Figure 13.

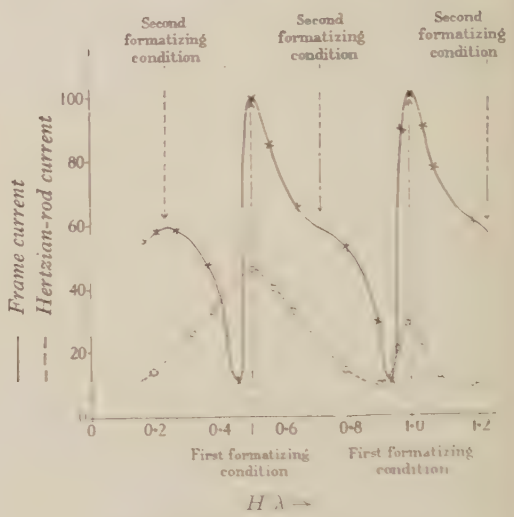


Figure 14. $\lambda = 5{\cdot}5$ m. $W/\lambda = 0{\cdot}36$.

difficulties inherent in these short-wave measurements. It is felt that perhaps the most crucial test of the foregoing theory is the marked increase in the radiation from frames the dimensions of which approximate to the theoretical values indicated by the co-ordinates of the points A and A'; whilst the fact that all the experimental

* Walmsley and Bewick in *G.P.O. Radio Report*, No. 211, of November 1931, record some tests which they carried out in order to verify certain preliminary conclusions published by Palmer and Honeyball in a letter to *Nature*, March 14, 1931. Walmsley and Bewick obtained maximum radiation from a square frame for which $H = W = 0{\cdot}50\,\lambda$ in one set of measurements and for which $H = W = 0{\cdot}47\,\lambda$ in a second set of measurements. They conclude: "From an examination of all the results it is evident that the size of frame is not very critical and that for all practical purposes there is no advantage in departing from the normal $\cdot 5\,\lambda$ square size of frame." They thus assume that their results are in accord with the old theory which stated that the frame current should be a maximum when the frame dimensions were equal to half a wave-length. Not only is this theory incorrect, but no radiation will occur if these conditions are exactly fulfilled. As their frame was only 4 ft. from the ground, their result is probably due to unsuspected reflected waves, and seems to be a further confirmation of the abnormal radiation predicted by the above theory for the square frame of sides equal to $0{\cdot}47\,\lambda$—a result which they actually obtained in one set of measurements. Walmsley and Bewick's largest frame was only $0{\cdot}56\,\lambda$ wide and hence they missed the large radiation which occurs when the frame-dimensions are given by the co-ordinates of the point A in figure 13.

points lie on the graphs of the equations of the first formatizing condition, but not on the graphs of the equation of the second formatizing condition, seems to give further support to the theory outlined in §§ 2, 4.

§ 6. CONCLUSIONS

(1) From the experiments described in § 3 it is concluded that variations in the dimensions of a tuned transmitting frame will appreciably affect the frame current and that the current will be a maximum only for certain critical values of the ratios of the dimensions of the frame to the length of the wave.

The critical frame-dimensions for the production of maximum current by a local oscillator are given approximately by the equations

$$\tan(a - \psi) = (a^2 - 1)/a \qquad \ldots\ldots (9)$$

and

$$\tan(a' + \psi) = (a'^2 - 1)/a' \qquad \ldots\ldots (10)$$

and also by equations (12) and (13) if waves reflected normally from the ground affect the frame current.

Less important frame-current maxima occur when

$$a + a' = n\pi \qquad \ldots\ldots (11).$$

This is the condition for the establishment of "permanent" oscillations or stationary waves round the frame with equally spaced nodes and antinodes.

(2) The experiments with the Hertzian-rod receiver, § 5, indicate that large radiation is obtained at points on the earth's surface in the plane of the frame when equations (9) and (10), or equations (12) and (13), are satisfied; but that the best radiation occurs when these equations, and also equations (1) and (2) with γ equal to 0, are simultaneously satisfied. That is to say, the best dimensions of a frame for horizontal radiation are:

$$H = 0{\cdot}40\lambda \text{ and } W = 1{\cdot}00\lambda \text{ (point } A, \text{ figure 13).}$$

With the effect of the reflected waves from the ground, additional dimensions for good radiation are:

$$H = W = 0{\cdot}47\lambda \text{ (point } A', \text{ figure 13).}$$

(3) No radiation perpendicular to the sides of the frame occurs when the frame-perimeter is an integral number of wave-lengths, i.e. when

$$a + a' = n\pi \qquad \ldots\ldots (11).$$

§ 7. ACKNOWLEDGMENTS

We should like to record our thanks to Mr R. Curry for assistance in carrying out this work, and also to Mr R. A. Watson-Watt who kindly loaned the short-wave oscillator used in the experiments.

DISCUSSION

See page 87.

621.396.674

THE CURRENT-DISTRIBUTION ROUND A SHORT-WAVE FRAME AERIAL

By Prof. L. S. PALMER, M.Sc., Ph.D., D. TAYLOR, M.Sc.
AND
ROY WITTY, B.Sc.

Received May 7, 1933, and in revised form August 8, 1933. Read October 20, 1933.

ABSTRACT. On the assumption that an initial oscillation in any part of a frame aerial is propagated round and round the frame an infinite number of times in both directions, with the consequent formation of stationary waves, a theoretical expression is obtained for the value of the current in any part of the frame. From the expression so obtained the positions and magnitudes of the current antinodes are deduced. The antinodes are fixed in space and are independent of the orientation of the frame to the wave-front. The current at any point of a frame consequently varies as the frame revolves in its own plane.

Experimental results are described which verify the conclusions concerning the positions of the antinodes and the effects produced by a revolving frame. Other experiments are in progress to test the deductions concerning the magnitudes of the currents at the antinodes.

§ 1. INTRODUCTION

In the previous paper* it was shown that the current produced in a tuned rectangular frame by a passing electromagnetic wave or by a local oscillator attained its maximum values when the ratio of the frame-dimensions to the wave-length was critically adjusted in accordance with the solutions of either of two equations.

In the case of a receiving frame actuated by an electromagnetic wave, the critical dimensions are given by

a, a' either
$$\left.\begin{aligned}\tan[a(1+\cos\gamma)-\psi] &= (a^2-1)/a\\ \tan[a'(1+\sin\gamma)+\psi] &= (a'^2-1)/a'\end{aligned}\right\} \quad \ldots\ldots(1),$$

or
$$a + a' = n\pi \quad \ldots\ldots(2),$$

n, W, H, λ, γ, ψ where $a = 2\pi W/\lambda$, $a' = 2\pi H/\lambda$, n is an integer, W is the frame-width, H is the frame-height, λ is the wave-length, γ is the angle of incidence of the wave-front to the frame and ψ is an arbitrary phase angle.

The corresponding equations for a transmitting frame actuated by a local oscillator are

either
$$\left.\begin{aligned}\tan(a+\psi) &= (a^2-1)/a\\ \tan(a'-\psi) &= (a'^2-1)/a'\end{aligned}\right\} \quad \ldots\ldots(3),$$

or
$$a + a' = n\pi \quad \ldots\ldots(2).$$

* Page 62 of this volume.

These sets of equations give what were called for convenience the first and second formatizing conditions for reception and transmission respectively.

When testing these equations experimentally it was found that the maximum currents which resulted when the frame was formatized were not uniformly distributed round the frame and that the current registered at any one point was not necessarily the same as that recorded elsewhere in the frame. Notwithstanding this spacial variation, the current at any given point did, in general, attain its maximum value when the frame-dimensions approximated to the solutions of either of the above equations.

In order to understand more fully the action of a short-wave frame aerial, it was thought desirable to investigate the current-distribution round a frame as the ratio of its dimensions to the wave-length varied.

It has already been stated in the previous paper that the second formatizing condition may be written

$$l = n\lambda,$$

where l is the frame-perimeter. From this it follows that stationary waves will be so formed that there will be $2n$ nodes and antinodes equally distributed round the frame, but there is no indication from this equation how these nodes and antinodes will be located with respect to the corners of the frame, or the position of the tuning condenser, or the position of the initial e.m.f. Neither is there any indication which of these three factors, if any, fixes the positions of the nodes and antinodes.

The first formatizing condition makes no stipulation as to the ratio of the frame-perimeter to the wave-length, and in general $l \neq n\lambda$ for the maximum currents given by equations (1) and (3). The problem thus resolves itself into a determination of the current-distribution round a short-wave frame aerial for the general case when the ratio of the frame-perimeter to the wave-length is not an integer. The current-distribution for the second formatizing condition given by equation (2), for which this ratio *is* an integer, will then be a special case of the more general solution.

§ 2. THEORETICAL CONSIDERATIONS

The action of a short-wave frame aerial. When an incident wave or a local oscillator actuates a frame aerial at a given point P, figure 1, the electric force causes a charge at P to oscillate to and fro at the wave frequency. The resulting disturbance travels round the frame with approximately the velocity of the wave in the surrounding medium. For convenience, the wave may be looked upon as starting off in one direction as a compression wave at the same time that it starts off in the opposite direction as a rarefaction wave. These two impulses will meet halfway round the frame at P' where they will both tend to accelerate a charge in the same direction; that is, their effects will reinforce each other, but will not necessarily be in phase with the disturbance at P' due directly to the passing wave, and a continuous disturbance will result at P' as long as the initial oscillation is maintained at P. These waves will continue round the frame and meet again at P where they

will again be in phase with each other, but not necessarily with the existing oscillation at P. In general, before they reach P, they will be met at Q_1 and Q_1', say, by the second impulses which started from P one period later and are now travelling away from P. At Q_1 the first compression impulse, say, from P will meet and reinforce the second rarefaction impulse from P, and vice versa at Q_1'. The time for the disturbances to travel from P to Q_1 by way of Q_1' and P' will be one period of the oscillation plus the time taken for the second impulse to travel from P to Q_1 directly. These two impulses at Q_1 (or Q_1') will be in phase and reinforce each other. Thus the disturbance at Q_1 (or Q_1') will reinforce when $l - x_1 = \lambda - x_1$, where l is the
x_1 frame-perimeter and x_1 is the distance PQ_1 measured the shortest way round the frame.

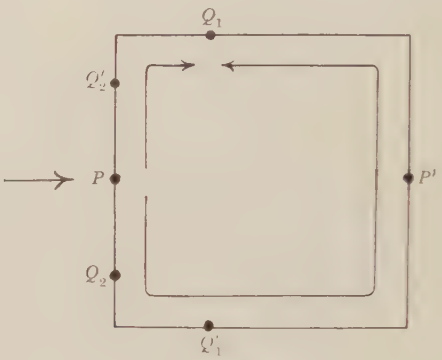


Figure 1.

In general, the two disturbances will have different amplitudes owing to the fact that the first disturbance will have travelled $(l - 2x_1)$ or λ round the frame farther than the second disturbance. These waves will continue to travel round and round in opposite directions and return to Q_1 (or Q_1') again in phase with each other, but with a further reduction in amplitude due to the loss during a complete circuit of the frame. Similarly, any two consecutive impulses at P will always reinforce each other at Q_1 and Q_1'.

By a similar argument it follows that alternate impulses at P will reinforce at
x_2 points Q_2 and Q_2' at a distance x_2 from P given by

$$l - x_2 = 2\lambda + x_2.$$

But with these current antinodes the amplitude of the disturbance will be smaller owing to the greater damping which will have resulted from the fact that the first impulse will have travelled a distance $(l - 2x_2)$ or 2λ farther than the third impulse.

In general, current antinodes with amplitudes decreasing with n will occur at
x_n distances x_n from P (or at distances $n\lambda/2$ from P') given by

$$x_n = \tfrac{1}{2}(l \pm n\lambda).$$

This equation may be written

$$a + a' - \delta_n = \pm n\pi \qquad \text{.....(4),}$$

δ_n where $\delta_n = 2\pi x_n/\lambda$ and the other symbols have the meanings indicated on p. 76.

A more complete treatment of the problem is as follows. Let the initial disturbance at P be $I_0 \sin \omega t$ and let the damping due to the path length of λ be η and that due to one circuit of the frame be ϵ, where η and ϵ are both less than unity. Then the resultant current at points Q_n and Q_n' due to the first reinforcement of the two impulses will be

η

ϵ

$$I_0 \sin \{\omega t - (\delta_n + 2n\pi)\} + \eta^n I_0 \sin \{\omega t - (\beta - \delta_n)\},$$
$$= I_0 \sqrt{\{1 + 2\eta^n \cos (\beta - 2\delta_n) + \eta^{2n}\}} . \sin (\omega t - \delta_n - \theta),$$

where $$\beta = 2\pi l/\lambda \text{ and } \theta = \tan^{-1} \frac{\eta^n \sin (\beta - 2\delta_n)}{1 + \eta^n \cos (\beta - 2\delta_n)}.$$

β, θ

These two waves will each complete a circuit of the frame and arrive back at Q_n (or Q_n') again with an additional phase-lag β and with a reduced amplitude ϵI_0.

Hence the resultant effect of these two waves will be

$$\epsilon I_0 \sqrt{\{1 + 2\eta^n \cos (\beta - 2\delta_n) + \eta^{2n}\}} . \sin (\omega t - \delta_n - \theta - \beta).$$

For the next circuit the resultant at Q_n (and Q_n') will be

$$\epsilon^2 I_0 \sqrt{\{1 + 2\eta^n \cos (\beta - 2\delta_n) + \eta^{2n}\}} . \sin (\omega t - \delta_n - \theta - 2\beta),$$

and the final result I at Q_n (and Q_n') is therefore

I

$$\sum_{Z=0}^{Z=\infty} \epsilon^Z I_0 \sqrt{\{1 + 2\eta^n \cos (\beta - 2\delta_n) + \eta^{2n}\}} . \sin (\omega t - \delta_n - \theta - Z\beta)$$
$$= I_0 \sqrt{\left\{\frac{1 + 2\eta^n \cos (\beta - 2\delta_n) + \eta^{2n}}{1 - 2\epsilon \cos \beta + \epsilon^2}\right\}} \sin (\omega t - \delta_n - \theta - \theta') \quad \ldots\ldots(5),$$

where $$\theta' = \tan^{-1} \frac{\epsilon . \sin \beta}{1 - \epsilon \cos \beta}.$$

θ'

Since the signs of the cosine terms in the numerator and denominator are plus and minus respectively, the amplitude of the current I is a maximum when either

$$\beta = 0 \text{ or } \pm 2n\pi,$$

i.e. $$a + a' = \pm n\pi \quad \ldots\ldots(2),$$

or

$$\beta - 2\delta_n = 0 \text{ or } \pm 2n\pi,$$

i.e. $$a + a' - \delta_n = \pm n\pi \quad \ldots\ldots(4).$$

Thus we recover equations (2) and (4) as before. It also follows from equation (5) that the values of the current maxima are not the same for the two conditions. When the former condition holds, then the amplitude of the current antinode is given by

$$I_{1(\max)} = \frac{I_0 \sqrt{\{1 + 2\eta^n \cos 2\delta_n + \eta^{2n}\}}}{1 - \epsilon} \quad \ldots\ldots(6),$$

I_1

which becomes infinite when $\epsilon = \eta = 1$ or damping is neglected.

When the second condition, equation (4), is fulfilled, then the maximum amplitude is

$$I_{2(\max)} = \frac{I_0 (1 + \eta^n)}{\sqrt{\{1 - 2\epsilon \cos 2\delta_n + \epsilon^2\}}} \quad \ldots\ldots(7),$$

I_2

which becomes $I_0 \operatorname{cosec} \delta_n$ when $\epsilon = \eta = 1$.

Some of the conclusions which follow from the interpretation of these equations will now be considered.

Some consequences. Deductions from the foregoing equations may conveniently be divided into those which concern the location of current antinodes round a frame and those which concern the magnitude of the current at the antinodes.

Since equation (2) is equation (4) with x_n equal to $-\frac{1}{2}n\lambda$, it follows that all the antinodes are spaced $\frac{1}{2}\lambda$ apart when the ratio of the frame-perimeter to the wave-length is an integer. Further, the first antinode, when $x = 0$, is at P, the point where the initial disturbance was produced. This will normally be the point of the frame nearest to the transmitter, or the point to which a local oscillator is coupled. Equation (2) has already been discussed at some length in previous papers. It has been called the "second formatizing condition." The justification for this may be seen from equation (6) which shows that, when the conditions given by equation (2) are fulfilled, the maximum amplitude of the current antinodes tends to be infinite, and is only limited by resistance damping. Thus the frame functions as a non-reactive and non-radiating circuit with ohmic resistance only.

Equation (4) has not been discussed hitherto and will therefore be considered in greater detail. There are, in particular, six conclusions concerning the frame current which call for consideration and experimental verification. The first four concern the location of the current antinodes and the last two deal with the current amplitudes. These six points are: (i) All the current antinodes are at fixed distances from P, which is a fixed point determined by the position of the initial disturbance. Hence the distribution of nodes and antinodes round a frame aerial is fixed in space and is independent of the orientation of the frame to the wave-front. (ii) The disposition of the current nodes and antinodes is symmetrical about the line PP', figure 1. (iii) Current antinodes always occur at P and P' irrespective of the wave-length. This follows because $x = \frac{1}{2}P$ when $n = 0$. (iv) Successive current antinodes are $\frac{1}{2}n\lambda$ from P' measured in both directions round the frame. (v) The amplitudes of the current at successive antinodes decrease to a minimum (I_0 if damping be neglected) as the distance x_n of the antinode from P increases, until this distance equals $\frac{1}{4}\lambda$, after which the current peak values gradually increase to a maximum (∞ if damping be neglected) when the value of x_n reaches $\frac{1}{2}\lambda$. This conclusion follows from equation (7), which shows that the amplitude is equal to I_0 cosec $[f(x_n)]$ when $\epsilon \quad \eta \quad 1$. (vi) For large values of n, the value of $I_{2(\max)}$, equation (7), decreases so that only the first few current antinodes may be readily detected in practice.

Finally, there is another consequence which follows indirectly from point (i) above. Since the nodes and antinodes are fixed in space, the current at any particular point in a frame aerial will vary as the frame revolves in its own plane, but the current will be constant for any point in the revolving frame when the point in question passes through any particular point in space. In other words, if a current-measuring instrument be fixed in a revolving frame aerial it will, apart from any other factors not here considered, record the current-values of the various nodes and antinodes as it passes through the points P, Q_2, Q_1', P', Q_1, Q_2', in succession. Also, if the instrument be fixed in space, the recorded current will be constant as

the frame revolves, the current value depending on the position of the instrument with respect to the transmitter or to the fixed points P and P'.

The experimental verification of the above six points and of the conclusions concerning a revolving frame would be a comparatively easy matter were it not for the fact that other disturbing factors may arise which tend to modify the experimental measurements. The chief of these factors is the current-variation caused by the fulfilment of the first formatizing conditions.

By employing a frame which is always deformatized from this standpoint—that is to say, by using a frame the dimensions of which do not satisfy equations (1) or (3)—it is possible to study the current-distribution which has been discussed above and to isolate one set of effects from the other.

§ 3. EXPERIMENTAL INVESTIGATION

Preliminary experiments have been carried out with a square frame of perimeter 12·24 m., and so arranged that the wire could slide over insulated pulleys fixed on the corners of the supporting wooden framework. Thus a vacuum thermojunction inserted at any point could either move round with the frame or remain fixed in space as the wooden framework moved, leaving the wire stationary.

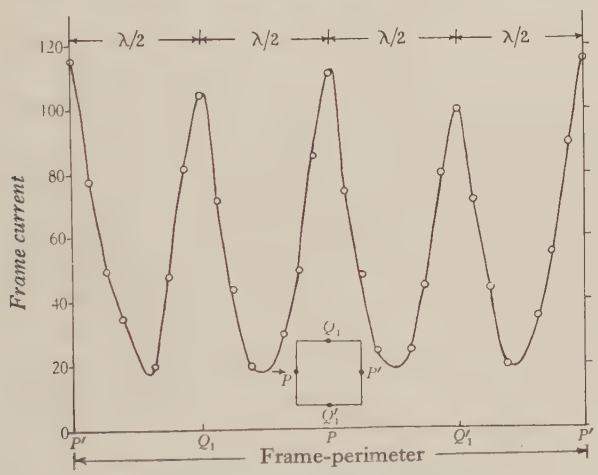


Figure 2. Fixed frame. $\lambda = 6{\cdot}12$ metres.

Some results are shown in figures 2 to 10. Figures 2 and 3 were obtained by moving a current-measuring instrument round a receiving frame which was fixed with one side nearest to the transmitter. Figure 2 is an illustration of the current-distribution when the frame-perimeter is an integral number of wave-lengths in accordance with equation (2), whilst figure 3 depicts the current-distribution when the ratio of the frame-perimeter to the wave-length is not an integer. Equation (4) is applicable to this case. These two figures show the fixed current antinodes at P and P' and the movement of the other antinodes Q_1 and Q_1' away from P' as the

wave-length increases. In each case the distances $P'Q_1$ and $P'Q_1'$ equal approximately $\frac{1}{2}\lambda$.

In figures 4 and 5 the circles show the theoretical positions, from equations (2) and (4), of the current antinodes, and the crosses show the experimental positions as determined from figures 2 and 3.

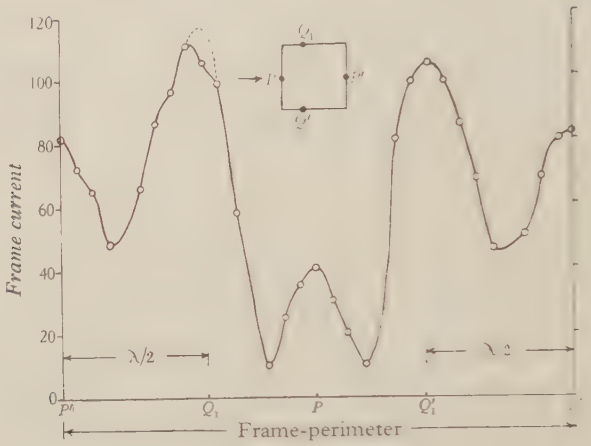


Figure 3. Fixed frame. $\lambda = 7{\cdot}3$ metres.

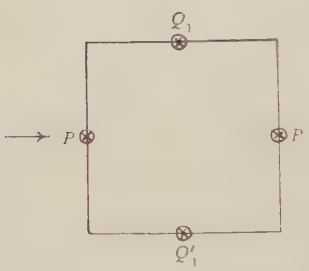


Figure 4. ⊙ Theory. × Experiment. $\lambda = 6{\cdot}12$ metres.

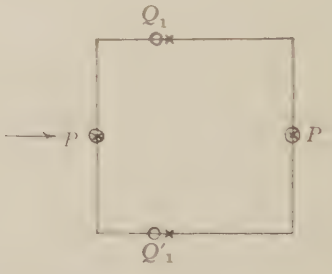


Figure 5. ⊙ Theory. × Experiment. $\lambda = 7{\cdot}3$ metres.

Figure 6 is a repetition of figure 3 with the frame oriented with one corner nearest to the transmitter. The comparison of this figure and figure 7 with figures 3 and 5 respectively shows the fixity of the current-distribution in space, irrespective of the orientation of the frame.

Figures 8, 9 and 10 were obtained with a revolving frame. Figure 8 is a reproduction of figure 2, but was obtained by rotating the frame together with the thermojunction. This again shows the fixity of the nodes and antinodes in space, but does not illustrate the current-variations predicted in a previous paper*, because the current-variations shown in figure 8 are not due to the frame becoming deformatized as it revolves. The frame remains formatized in accordance with the

* *Proc. R.S.* A 136, 193–209 (1932).

second formatizing condition, equation (2), and is always deformatized from the point of view of the requirements of equation (1). These current-variations are only due to the fixity of the positions of the current nodes and antinodes in space.

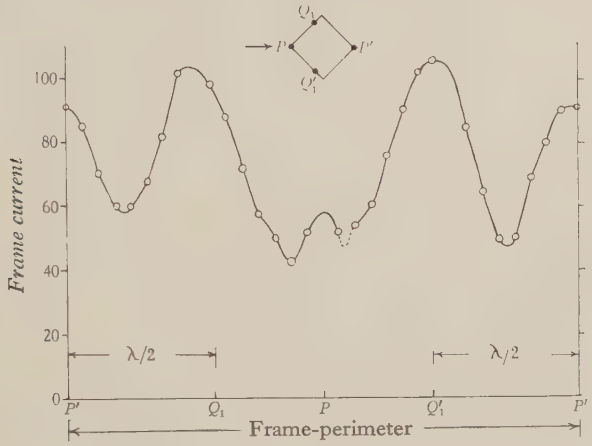


Figure 6. Fixed frame. $\lambda = 7{\cdot}3$ metres.

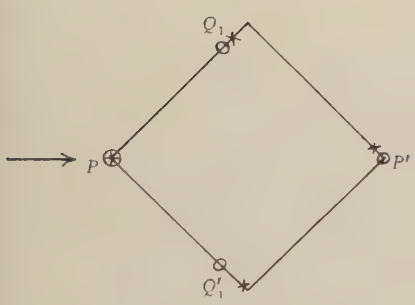


Figure 7. ⊙ Theory. × Experiment. $\lambda = 7{\cdot}3$ metres.

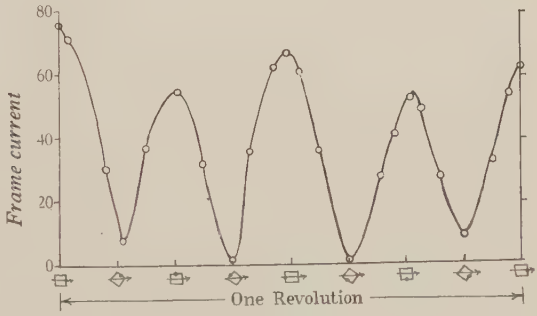


Figure 8. Revolving frame. $\lambda = 6{\cdot}12$ metres.

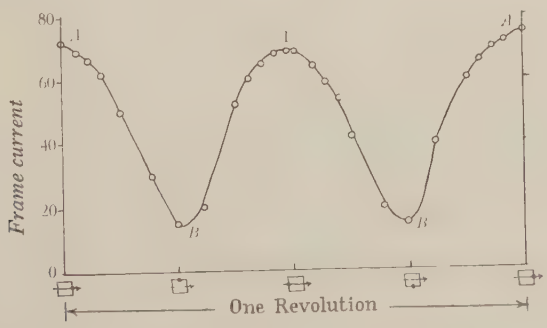


Figure 9. Revolving frame. $\lambda = 12{\cdot}2$ metres.

Figure 9 is similar to figure 8, but the ratio of the perimeter to the wave-length is unity. The only antinodes are those at P and P'.

Finally, figure 10 confirms the conclusion that if a frame be rotated and the current-measuring instrument be fixed in space, then the measured current will remain constant and record a value depending on the distance of the instrument from the fixed points P and P'. The values of A and B in figure 10 refer respectively to the points A and B in figure 9, which indicate where the instrument was fixed.

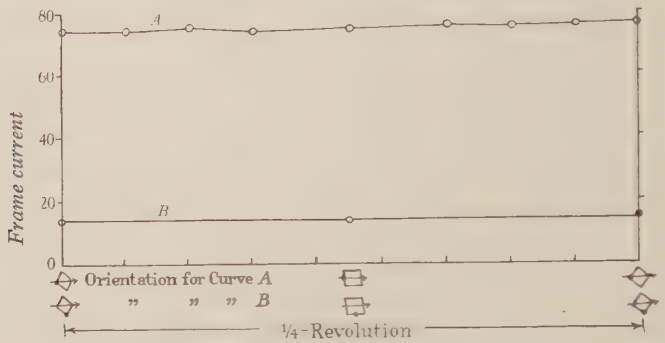


Figure 10. Revolving frame with instrument fixed in space. $\lambda = 12{\cdot}2$ metres.

§ 4. CONCLUSIONS

The above preliminary experiments confirm the first four of the six points discussed in § 2 above, but there are not yet sufficient data to verify points (v) and (vi) concerning the magnitudes of the amplitude-variations. Further experiments to this end are now in progress. The experiments with the revolving frame are also in accordance with the conclusions deduced at the end of § 2 on pp. 80 and 81.

Hence it is concluded that, as long as the ratios of the frame-dimensions to the wave-length do not satisfy equations (1) or (3), equation (5) gives the position, and possibly the magnitudes, of the current antinodes round a short-wave frame aerial.

§ 5. ACKNOWLEDGMENT

We are pleased to record our thanks to Mr R. Curry, B.Sc., D.I.C., for assistance in carrying out the experiments.

DISCUSSION

See page 87.

621.396.674

DEMONSTRATION

Some experiments with ultra-short wireless waves in water. *Demonstration given on October* 20, 1933 by Prof. L. S. PALMER, D. TAYLOR and ROY WITTY *to illustrate papers on pages* 62 *and* 76 *entitled respectively "The action of a tuned rectangular frame aerial when transmitting short waves" and "The current-distribution round a short-wave frame aerial."*

In order to demonstrate the effects discussed in the two papers referred to above it is necessary to employ frame aerials comparable in dimensions with the length of the wireless wave in use. This necessity introduces two difficulties if the demonstrations are to be carried out in a lecture room. Firstly, the frames must be of such dimensions that they can be manipulated easily on a demonstration bench, and secondly, the radiated power from the short-wave transmitter must be sufficient to operate flash-lamp bulbs or other visible current indicators which may be inserted at various points round the frames where it is desirable to show the presence or otherwise of current antinodes. If the size of the frames be of the order of 10 or 20 cm., then the available power from any simple transmitter such as a Barkhausen valve oscillator operating on such short wave-lengths is insufficient. With wave-lengths greater than 3 m. the radiation can be adequate, but in this case the frames must be inconveniently large and therefore difficult to manipulate in a limited space.

It was found that both these difficulties could be overcome by using an ordinary valve oscillator operating on 3 m. or 4 m. (say) in air but coupled to a radiating half-wave Hertzian dipole immersed in a tank of water. Thus the radiated wave-length is reduced in the ratio of 1 to $1/\sqrt{K}$, where K is the dielectric constant of water for these frequencies. K is about 80, so that the normal wave-length range of 3 to 4 m. in air is only 33·6 to 44·7 cm. in water. With this method the available power is found to be ample for actuating small receiving frames of the order of 10 to 20 cm. in linear dimensions when immersed in the glass tank of water containing the radiating dipole. It is convenient to transmit vertically downwards by having the dipole arranged horizontally about a quarter of a wave-length below the surface of the water.

Figure 1 is a diagrammatic sketch of the actual apparatus in which the four essential parts of the transmitter are the power unit A, the valve oscillator B, the half-wave transmission line C, and the radiating dipole D. In addition there is the tank of water E containing the dipole D and the receiving frame F. The motor G revolves the frame in its own plane by suitable gearing.

The circuit diagram is shown in figure 2 in which the reference letters correspond with those in figure 1.

To show the current-distribution round different frame aerials when actuated by waves of different lengths, it is convenient to use circular frames with perimeters ranging from 1 to 2 wave-lengths, that is from about 40 to 80 cm. With a frame, the perimeter of which is equal to one wave-length, current antinodes occur at those

points of the frame which are nearest to and most remote from the dipole D. These are the antinodes marked P and P' on page 78 of the paper on "The current-distribution round a short-wave frame aerial." To indicate the positions of these antinodes (which are fixed in position) a flash-lamp bulb is inserted in the frame. The frame is mounted on a wooden support which can be revolved by the motor G fixed outside the tank. In this way, as the frame revolves in its own plane, so the flash-lamp bulb passes in succession through the fixed points P and P', and in so doing it is momentarily lighted up. By the use of a frame with a perimeter greater than one wave-length, the positions of the additional antinodes Q and Q' can be observed. Also, if the wave-length be gradually increased, the resulting gradual motion of these antinodes away from P' can be detected by the variation in the positions at which the lamp becomes lighted.

Figure 1

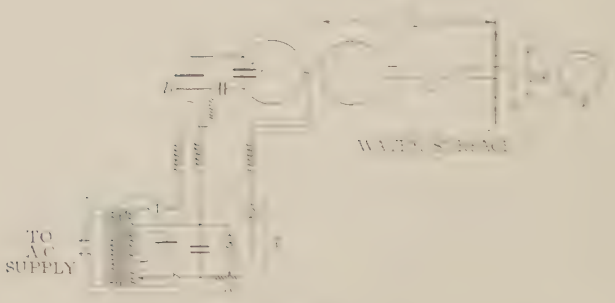


Figure 2

With a square or rectangular frame the same results are obtained whatever the positions of the lamp in the frame as long as the frame-dimensions compared with the wave-length are such that the frame remains deformatized* as it revolves. This can best be shown by taking a square frame with two flash-lamp bulbs in it, one in the middle of one side of the frame and one at a corner. Then, as each bulb passes in turn through the positions of the antinodes P or P' it lights up, and the intensity of the light from each bulb is the same.

To show the effect produced when a transmitting frame is formatized in accordance with the conditions given by equations (9) and (10)† is not practicable with this apparatus; but the effect of formatizing a receiving frame in accordance with

* *Loc. cit.* p. 62.

† *Loc. cit.* p. 66.

equations (1) and (2) may be demonstrated by using a square frame with two flash-lamp bulbs inserted where indicated above. The frame-dimensions compared with the wave-length must be such that the frame is formatized when one side is parallel to the dipole D, i.e. when the frame is in the square position for which γ in the above equations (1) and (2) is equal to zero. Then, as the frame revolves the lamp in the side of the frame lights up more brightly than the lamp in the corner. This effect is reversed if the frame be of such dimensions that it is formatized in the diamond position with its sides at 45° to the dipole, i.e. if $\gamma = 45°$ in equations (1) and (2). Thus although the antinodes remain fixed in space they do not remain constant in amplitude if the frame becomes alternately formatized and deformatized as it revolves.

For a wave-length of 3·8 m. in air or 42·5 cm. in water the sides of a square frame must be 20 cm. long for the frame to be formatized in the square position, whilst the sides of the frame must be only 16·5 cm. long for the frame to be formatized in the diamond position with this particular wave-length. These values may be deduced from figure 3 of a previous paper*.

These three experiments with square frames serve as a useful qualitative test of the theory which has been advanced to explain the anomalous effects obtained when rectangular frames are used with short waves.

By the use of the apparatus described above together with a small Hertzian dipole receiver with a flash-lamp bulb at its centre, the polarization of wireless waves and the changes produced on reflection from various surfaces may be readily demonstrated. With a sufficiently large tank, interference phenomena can be observed; and from the positions of the nodes and antinodes on Lecher wires in various liquids the dielectric constants and their variation with frequency may be measured. Because it is possible with this arrangement to illuminate flash-lamp bulbs, the apparatus is particularly convenient for demonstrating most of the properties of electromagnetic waves to large audiences.

DISCUSSION OF THE PRECEDING TWO PAPERS AND DEMONSTRATION

Dr R. L. Smith-Rose. The subject of frame aerials for the transmission and reception of short electric waves is one that is of considerable importance and interest to those concerned with radio communication. For ordinary frame-dimensions and medium wave-lengths the justifiable assumption is usually made that the ratio of height or width to wave-length is small. When, however, the wave-length is reduced to such an extent that this ratio is no longer small compared with unity, then certain effects outlined by Prof. Palmer in his previous papers become important and must be taken into account. Some of us have thought that the discussion of the subject given in these earlier papers was somewhat unsatisfactory, and we are therefore glad to see that the matter is being pursued further.

* L. S. Palmer, *Proc. R.S.* A 136, 199 (1932).

One of the points that has been cleared up is the question of the current-distribution around the frame, and (as the experiments clearly demonstrate) it is not sufficient to refer to the current in a frame without specifying the point at which the current is being measured, in addition to the wave-length and frame-dimensions. It would be interesting to know why the point P was selected at the centre of a vertical side in the theoretical analysis of the reception by a frame of an induced e.m.f. Presumably the incident electric force must be parallel to the side containing P. Is the location of P decided merely from conditions of symmetry? Since the frames do not include any tuning-condensers the current-distribution might be deduced from that in a transmission line or a Lecher wire system.

It would appear that the current-distribution in some of the cases illustrated in the paper could have been deduced from the suggestions of current-distribution in open wire aerials put forward by Korshenewsky*. Incidentally since, I believe, Korshenewsky's ideas have never been verified experimentally, I would suggest to Prof. Palmer that the technique he has developed for the study of current-distribution might be applied to the distribution on open-wire aerials at short wave-lengths. It is to be presumed that the current at points P and P' is in opposite phase to that at Q_1 and Q_1' and this might have been shown in the diagrams by plotting with negative ordinates. In this case the curves would not have been symmetrical about the zero axis, and it would be interesting to know the explanation of the asymmetry in, say, figure 2.

In the first paper the authors have discussed the optimum conditions of a frame when used as a transmitting aerial on very short wave-lengths. Since it has already been shown that the mere measurement of current at a particular point in a frame has little meaning by itself, the important part of this paper is that dealing with conditions of maximum radiation. From this viewpoint the criticism of the work of Walmsley and Bewick at the foot of p. 74 appears to be scarcely justified, for the following reason. In figure 12 the optimum size of frame for maximum radiation is that corresponding to a height of $0{\cdot}4\lambda$ and a width of $0{\cdot}5\lambda$, while if the height is increased to $0{\cdot}5\lambda$ the radiation field is only reduced by about 25 per cent. Now the optimum sizes of frames found by Walmsley and Bewick are $0{\cdot}47\lambda$ and $0{\cdot}5\lambda$ square in two experiments respectively; yet the authors state in their footnote on p. 74 that no radiation will occur if these conditions are exactly fulfilled. The explanation of the result as being due to the effect of the waves reflected from the earth seems scarcely sufficient, and if it is correct it must always be considered in connexion with such results as those illustrated in figures 4 and 12. Furthermore, frames of the dimensions under consideration will have markedly directional radiation characteristics and these must be taken into account in comparing the measured radiation from frames of different dimensions.

From the radio-communication engineer's point of view the real point at issue here is to know the relative advantages of the so-called "formatized" frame and a suitable antenna array occupying the same spatial dimensions and supplied with the

* N. von Korshenewsky, *Z. f. Technische Physik*, **10**, 604 (1929); see also F. M. Colebrook, *J. Inst. E.E.* **71**, 235 (1932).

same input power. Does the experience of the authors throw any light on this aspect of the communication problem?

Dr L. E. C. Hughes. The method of treating antennae of dimensions comparable with the wave-length as given by the authors is particularly interesting and likely to be of great value when applied to the more complicated directional antennae used in practice. A new type of antenna, called the diamond or rhomboid, is being tried out on both sides of the Atlantic. It consists of four wires in a horizontal plane, with the angles of maximum radiation (and also reception) so arranged that the effects are additive along the direction of the longer diagonal and tend to cancel perpendicularly for the four wires taken together. This is based on the principle that the angle of maximum radiation from a non-resonating wire appears to reach a limit of about 70° when the wire is more than about 6 wave-lengths long. A feature of the arrangement is that the termination at the far end of the diamond, the two sides being fed in opposite phases, renders the antenna aperiodic for a wide band of frequencies. The exact mode of operation of this antenna seems to be obscure, but an extension of the work contained in the paper promises a solution.

Authors' reply. Dr Smith-Rose has raised several interesting points. Because the work is necessarily and unavoidably incomplete at the moment, we realize that no such simple theory as that now set forth can completely explain the action of short-wave frame aerials, but it seemed desirable to publish the work as far as it has gone together with the theory as far as it has been developed. In this respect the present communication may be, we fear, almost as unsatisfactory as the earlier publications.

The point P is the point nearest to the radiating source and therefore is the pole of the spherical wave-front when incident on the frame. The vertical component of the electric vector is therefore a maximum at P and not uniform. In the case of a point source located near the centre of the frame, the point P can be made to travel round the frame in a manner which depends on which part of the frame is nearest to the point source. We have not been able to obtain experimentally a perfectly symmetrical case in which an ideal point source is exactly in the centre of a circular or square frame. The point P is also independent of the position of any tuning-condenser which may be in the frame.

Concerning Korshenewsky's work and the recent work of Dr Smith-Rose and his collaborators at the National Physical Laboratory, we originally attacked the current-distribution question from the transmission-line standpoint but felt that the necessity of bending the wires to form a frame led to difficulties at the point of contact owing to the terminal impedance of the lines, and, as the method seemed somewhat artificial when applied to closed frames, we have, at the moment, abandoned it.

The reason for the lack of symmetry about the zero axis in the current-graphs in the second paper, when alternate peaks measured from P' are plotted with negative ordinates, is partly due to inability to measure small currents with a square-law instrument. Probably our minimum current-values were in reality zero in the cases

where the ratio of the wave-length to the frame-perimeter was approximately an integer, whilst the presence of minor Q_1 and Q_1' antinodes round the frame when this ratio was not an integer would account for small currents at the positions of the major current nodes.

Concerning our criticism of Walmsley and Bewick's paper, our objection is not to their experimental values, which approximate closely to our own, but to their conclusions. We do not believe that their maximum currents obtained for $H = W = 0{\cdot}47\lambda$ and $0{\cdot}5\lambda$ are due to their frame being "the normal $0{\cdot}5\lambda$ square size of frame." We think that the old theory which leads to this conclusion regarding the frame-width is erroneous. When a frame of these dimensions is raised high above the ground, both the frame current and the horizontal radiation from the frame vary and diminish, showing that the ground is responsible for the effect described; whilst, if a frame of dimensions $W = 1{\cdot}0\lambda$ and $H = 0{\cdot}4\lambda$ be similarly raised, its circulating current and radiation properties remain practically unaltered. The large permanent oscillations which are set up in a frame for which $H = W = 0{\cdot}5\lambda$ preclude a large radiation resistance. In our experiments we have been unable to detect any appreciable horizontal radiation when the frame-dimensions were such that permanent oscillations were set up. So far we have only considered horizontal radiation perpendicular to the sides and in the plane of the frame (i.e. the case in which $\gamma = 0$).

We are in complete agreement with Dr Smith-Rose concerning the desirability of knowing the relative efficiency of a formatized frame and an antenna array; but although we believe a formatized frame may be better than a half-wave dipole with a suitable reflector, we have not yet had time to carry out any experiments on relative efficiencies.

In reply to Dr Hughes: the work to which he refers was, we believe, first carried out in America by E. Bruce*. Mr R. Curry at Hull is now investigating the problem of formatizing triangular frames for reception and transmission, but we have met with considerable difficulty in applying the theory of this paper to frame-wires which are not parallel. We believe that the solution of the triangular frame may throw considerable light on the diamond aerials to which Dr Hughes refers, but the problems are different in that the frames we are studying are closed circuits and do not radiate and receive by virtue of critically terminated open ends.

* *Proc. Inst. Radio Eng.* **19**, 1406 (1931).

551.51.053.5:621.396.11

THE DIURNAL VARIATION OF THE INTENSITY OF WIRELESS WAVES REFLECTED FROM THE IONOSPHERE

By F. W. G. WHITE, M.Sc.
(Halley Stewart Laboratory, King's College, London.)

Received May 15, 1933. *Read November* 3, 1933.

ABSTRACT. An account is given of the diurnal variation of the relative intensity of waves returned from the ionized regions of the upper atmosphere as observed over the early morning period from about 2 a.m. until about 9 a.m. The relative intensity is taken as the ratio of the intensity of the downcoming wave to that of the ground wave. The influence on the intensity of electron-limitation and of absorption-limitation is discussed, in the light of theoretical ionization curves given by Chapman. The experimental observations described show that the magneto-ionic doubling of the echo, which has been observed by Appleton and Builder for the F region, occurs also for the E region. The doublet echo for this region is observed only for a short period of the morning, owing to the fact that the extraordinary ray is very soon totally absorbed.

§ 1. INTRODUCTION

ALTHOUGH within the last few years much work has been done in the study of the equivalent height of reflection of wireless waves from the upper atmosphere and in the determination of their polarization on arrival at the ground, the study of the efficiency of reflection has not received corresponding attention. In an early discussion of the factors controlling the intensity of downcoming waves Appleton pointed out that there were two processes that would tend to make the reflection coefficient less than unity. In the first place, owing to collisional friction, the waves may be absorbed in passing through one of the ionized regions, thus suffering what Appleton termed "absorption limitation." Secondly, there may be an insufficient electron-density to cause an appreciable return of the energy, in which case the impaired reflection is due to "electron-limitation." The object of this paper is to consider the relative influence of these two factors in certain practical cases.

On the theoretical side we know, from the simple Lorentz theory of dispersion, that when no impressed magnetic field is operative the effect of absorption-limitation can be expressed in terms of an absorption coefficient k given by

$$k = \frac{\nu}{2c\mu} \cdot \frac{4\pi N e^2}{m(p^2 + \nu^2)} \qquad \text{.....(1),} \qquad k$$

in such a way that the ratio of the intensity of the emergent wave to that of the incident wave is given by

$$\frac{E}{E_0} = \epsilon^{-\int k ds} \qquad E, E_0, \epsilon, s$$

μ The refractive index μ in the above expression (1) is given by

$$\mu^2 = 1 - \frac{4\pi N e^2}{m(p^2 + \nu^2)} \qquad \ldots\ldots(2),$$

where

c c is the velocity of light in free space;

ν ν the collisional frequency of the electrons in the medium;

N N the number of electrons per cm^3;

e, m e, m the charge and mass of the electron;

p p the angular frequency of the waves.

In this paper the absorption caused by the electronic component of the ionized regions will be discussed. It has definitely been established by Appleton that in the upper or F region refraction and absorption are caused chiefly by electrons, and it seems probable that this is also true for the E region.

The above relationships are not valid if μ changes rapidly in a distance comparable with the wave-length of the waves in the medium. As a typical example of true absorption we have the case of a wave, reflected from the upper or F region, being absorbed in the lower or E region through which it must pass.

In terms of Appleton's analysis the waves will be completely refracted downwards when they reach a point where there are sufficient electrons to reduce the refractive index to zero for vertical incidence. It is not, however, a sufficient condition for the complete return of the energy of the waves that μ should be reduced to zero at one level in the ionized region. Hartree has shown for several special cases that the coefficient of reflection depends also upon the total thickness of the region, measured in terms of the free-space wave-length of the waves. The case considered by him which most closely approximates to that of the E region is that

μ_m in which there is a linear decrease of μ^2 to a minimum value μ_m^2 as the height above the earth's surface increases, with a similar increase to $\mu^2 = 1$ for a further increase in height. Excluding absorption and taking the case of the normal incidence of plane waves, the region will reflect only a quarter of the energy for $\mu_m^2 = 0$, even when the thickness is very great compared with the wave-length. A small amount of energy will be reflected before there are sufficient electrons to reduce the refractive index to zero, that is when $\mu_m^2 = 0$. A reflection coefficient nearly equal to unity will be obtained only when there are more than sufficient electrons to reduce μ_m^2 to zero – that is when $\mu_m^2 = -0{\cdot}1$ in terms of equation (2)*. If we suppose that in the earth's atmosphere the variation of the refractive index with height is similar to that chosen by Hartree and that the thickness is great compared with the wave-length, then, as the ionization increases with time of day, a reflected wave of small intensity will be observed before the minimum value of μ^2 is zero. When it is just zero the amplitude of the reflected wave will be about 50 per cent

* It should be noticed that in terms of the original analysis the right-hand side of equation (2) is equal to $c^2k^2p^2$ for negative values if we keep μ^2 only for the positive values in this equation. The use of negative values of μ^2 is a convenient way of expressing the fact that the real value of the refractive index is zero in a certain thickness of the refracting regions.

of that of the incident wave and we shall expect perfect reflection only when μ_m^2 becomes negative and equal to about $-0{\cdot}1$. The amplitude of the reflected wave will be small, on account of this phenomenon of electron-limitation, when during the diurnal variation of the ionization either layer just begins or just ceases to reflect the waves. It must be remembered that electron-limitation and absorption may be effective at the same time.

§ 2. THEORETICAL DISCUSSION

In order to obtain an approximate idea of the cause of the diurnal variation of the relative intensity of the downcoming waves, a possible case is discussed below for which use may be made of the theoretical curves given by Chapman* for the variation of the ionization with height in the atmosphere at various times of the day for latitude 60°. This discussion is confined to the case of the early morning period in summer, when, after sunrise, the ionization begins to increase. The early

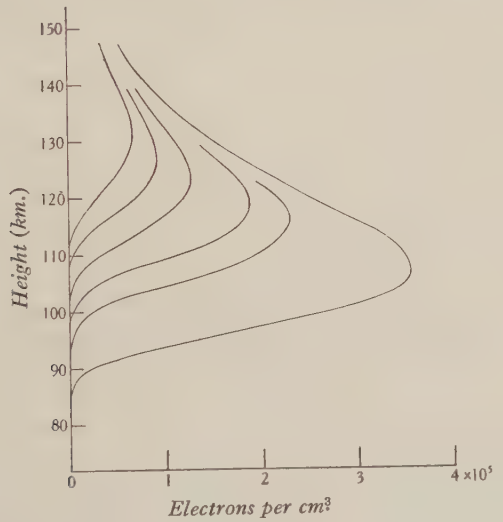


Figure 1. Assumed ionization at 3.36, 4.00, 4.22, 5.00 and 5.54 a.m. and noon.

morning period is of special interest because of the rapid change that takes place in the ionization in consequence of the incidence of the effective solar rays on the atmosphere. The curves of figure 1 show the variation of ionization with height for a midsummer morning. Although they are based on Chapman's curves they differ from them in several minor respects which will be discussed later. These curves correspond to the variation of ionization with height in the E region, the actual value of the number of electrons per cm^3 being obtained by assuming that Chapman's value $n/n_0 = 1$ is equivalent to 4×10^5 electrons per cm^3, the average value given by Appleton and Naismith† for summer in approximately this latitude. The height

* *Proc. Phys. Soc.* **43**, 26 (1931).
† *Proc. R.S.* A **137**, 36 (1932).

at which the ionization occurs has been adjusted to correspond to that of the E region to a first approximation.

In order to calculate the absorption index, the variation of the collisional frequency (ν) of the electrons with height must be known. The values used in this paper are those given by Pedersen*. This variation is shown in figure 2.

Suppose that a spherical wave is sent upwards from a point on the earth's surface and that the downcoming waves are received at a point near the transmitter, so that we are dealing with a case of substantially vertical incidence. The wave-length is chosen to be 100 metres, a wave-length frequently used in the experiments. Owing to the variation of the time of sunrise with height above the ground, region F will be first irradiated by the sun and the ionization will begin to increase there before it does so in the E region. Since the rate of recombination in

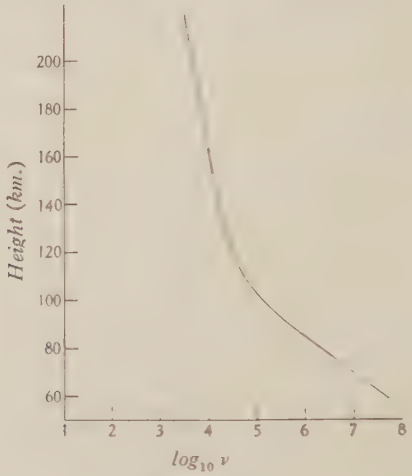


Figure 2.

region F is probably low, the ionization will build up fairly rapidly and it is reasonable to suppose that region F will be a good reflector for 100-metre waves shortly before sunrise at the E region. In the following work it is convenient to compare the intensity of the downcoming wave (denoted by F) from the F region or (denoted by E) from the E region with the intensity of the ground wave (denoted by G) at the receiver. If we assume that at 2.45 a.m. the F region is a good reflector and also that there is little absorption in the E region, then at this time we may take the ratio F/G to be unity. The ratio F/G or E/G at other times of day may then be compared with F/G at 2.45 a.m.

Appleton and Ratcliffe have previously shown that the total absorption suffered by the waves in their atmospheric path may be found by measuring the "reflection coefficient of the ionized regions." This quantity ρ is so defined that it is equal to $\alpha F d_1 / G d_0$, where F and G are as defined above, while d_1 and d_0 are the distances

* *The Propagation of Radio Waves* (Copenhagen). The values of ν used above are those given by Pedersen for the atmosphere which he has called "atmosphere F."

travelled by the atmospheric and ground waves respectively. The constant α depends upon the radiation characteristics of the transmitting aerial and upon the attenuation of the ground wave. In the experimental work described below α could not be determined, so that relative values of the ratio F/G or of E/G have been measured.

At later times of day, irradiation of the E region will cause the ionization shown in figure 1. The first effect of the increase of ionization in the E region will be to interpose an absorbing region in the path of the waves reflected from the F region. Since such waves pass twice through this absorbing zone, the total absorption will be $2\int kds$, the integral being evaluated for the total thickness of the E region. At

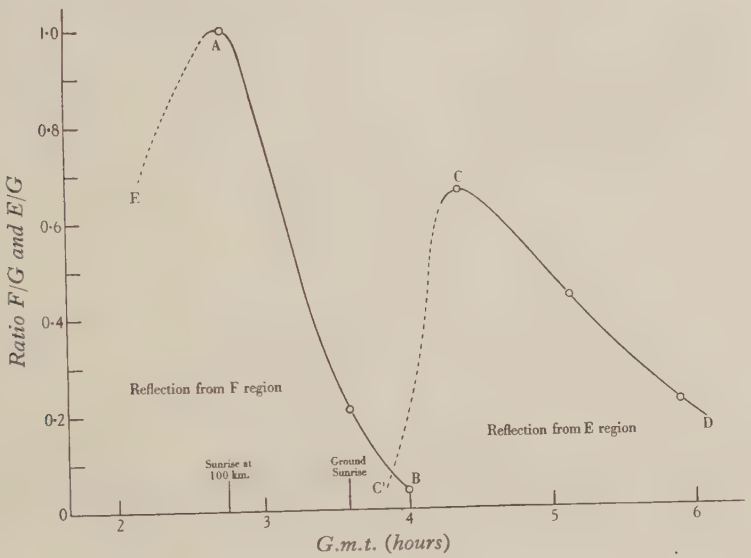


Figure 3.

times later than 2.45 a.m., therefore, the ratio F/G will be reduced from unity to $\epsilon^{-2\int kds}$. This ratio has been calculated by means of the formula

$$k = \frac{\nu}{2\mu c}\frac{4\pi Ne^2}{mp^2} \quad (p^2 \gg \nu^2)$$

by graphical methods for the times 3.36 and 4.00 a.m. The results of the calculation are shown in the curve of figure 3 from A to B.

The number N_0 of electrons required to reduce the refractive index to zero, for 100-metre waves, is from equation (2) given by

$$N_0 = \frac{mp^2}{4\pi e^2} = 1{\cdot}115 \times 10^5 \text{ per cm}^3 \quad (p^2 \gg \nu^2).$$

By 4.22 a.m. there are sufficient electrons in the E region to cause the return of these waves, so that at some time between 4.00 and 4.22 a.m. the waves will be intercepted by the E region and returned from it. Since we are considering spherical waves, account must be taken of the fact that when they are returned from the

E region the distance of travel is approximately half that for the F region. Thus at 4.22 a.m. we shall have $E/G = 2$ for no absorption. Taking into account the absorption in the lower half of the E region the ratio becomes $2\epsilon^{-2\int k ds}$. In considering the absorption of the waves in the lower part of the reflecting E region we must remember that the waves travel into the region until they reach a point where there are sufficient electrons to reduce the refractive index to zero. At this point the index of absorption k becomes infinite, and the question arises whether this simple theory is inadequate in that it gives infinite absorption for such a reflection. That it does not do so can be seen by examining the absorption for the two cases in which the ionization increases linearly and parabolically with height.

h_0 Suppose that at a height h_0 above the earth's surface the number of electrons per cubic centimetre begins to increase and does so according to the equation $N = N_0 y/y_0$, where y is the distance measured above h_0. For N_0 it is best to choose the value $N_0 = mp^2/4\pi e^2$ so that y_0 is the actual distance travelled by the waves in the ionized region. Then

y

$$\mu^2 = 1 - y/y_0.$$

To evaluate $2\int k ds$ we have therefore

$$2\int k ds = \frac{1}{c}\int_0^{y_0} \nu \frac{1-\mu^2}{\mu}\, dy$$

$$= -\frac{2y_0}{c}\int_1^0 \nu\,(1-\mu^2)\, d\mu \qquad \text{.....(3),}$$

where ν is a function of y and therefore of μ. Although the index k becomes infinite at $y = y_0$ the total absorption is finite.

If, above h_0, N increases according to the equation $N = N_0 y^2/y_0^2$, where $N_0 = mp^2/4\pi e^2$ and y_0 is the actual distance travelled by the waves in the region as before, then we have $\mu^2 = 1 - y^2/y_0^2$ and

$$2\int k ds = \frac{1}{c}\int_0^{y_0} \nu \frac{1-\mu^2}{\mu}\, dy$$

$$= -\frac{y_0}{c}\int_1^0 \nu\,(1-\mu^2)^{\frac{1}{2}}\, d\mu \qquad \text{.....(4).}$$

Thus in both cases the total absorption is finite.

An examination of the curves of figure 1 shows that the ionization follows very closely a parabolic distribution up to the value $N_0 = 1{\cdot}115 \times 10^5$ electrons per cm.^{3} The absorption of the waves in the lower part of the E region has therefore been calculated by the aid of equation (4), the integral being evaluated graphically for the times 4.22, 5.00, 5.54 a.m. and noon. The corresponding values for $E/G = 2\epsilon^{-2\int k ds}$ are shown in figure 3 from C to D.

There is at present no theoretical analysis describing the effect of the thickness of the refracting region on the coefficient of reflection for spherical waves. We may, however, roughly estimate the effect of electron-limitation on the relative intensity of waves refracted from the E region by reference to the results of Hartree for plane waves. Since $\mu^2 \propto N$ we can see from figure 1 that μ^2 will decrease to a

minimum (at the maximum value of N) and then increase again as the height is increased, in a manner similar to the triangular variation of the square of the refractive index with height chosen by Hartree. The total thickness of the region is always great compared with the wave-length. We should expect therefore that at 4.22 a.m. the effect of electron-limitation will be negligible but that between this time and 4.00 a.m. it will be of increasing importance. It is probable therefore that the relative intensity of the waves returned from the E region will increase from a small value at about 4.00 a.m. to the value shown in figure 3 at 4.22 a.m. This increase has been indicated by the dotted line in that figure (C'–C).

Figure 3 gives now the diurnal variation of the relative intensity of the waves from a time just before sunrise until well into the morning. The waves are at first returned from the F region with maximum relative intensity. Their intensity decreases rapidly as time advances owing to absorption in the E region where the ionization is building up. At about 4.00 a.m. there is nearly sufficient ionization in the E region to cause the waves to be returned from that region. The E reflection increases in intensity owing to the disappearance of electron-limitation until 4.22 a.m., when the waves from the E region have their maximum relative intensity. After this time the relative intensity of these waves decreases owing to absorption in the lower part of the E region itself. The above discussion has attributed all the absorption to the E region. It is perhaps permissible to do this, for the collisional frequency of the electrons in the E region is of the order of 100 times that for the F region, so that for the same ionization the absorption index will be also about 100 times greater.

We must now try to decide on the cause of the variations in the intensity of the waves reflected from F region before 2.45 a.m. After sunset on the previous day the ionization in regions E and F begins to decrease, with the result that the reflection of a 100-metre wave from region E and ultimately from region F may cease and no reflected wave be observed at the ground. When the influence of the sun is once again felt in region F the ionization increases again, and taking into account the electron-limitation effect in that region, we should observe first a small signal, which increases in intensity as the region becomes more perfectly reflecting. This has been indicated in the dotted portion of figure 3 from E to A. Reflection from the F region will cease during the night only if the maximum number of electrons decreases to a value less than the critical value $N_0 = mp^2/4\pi e^2$. If reflection from region F persists throughout the night we should expect the waves to have a steadily increasing intensity, owing to the continual removal of the absorbing E region.

§ 3. THE EARTH'S MAGNETIC FIELD

From the Lorentz theory we know that if there is an impressed magnetic field operative in the region of ionization, then the motion of the electrons is influenced by it in such a way that the refractive index and the absorption coefficient are functions of the intensity of the field. The analysis of the magneto-ionic theory of

propagation of a wave, in any direction with respect to the applied field, has been given by Appleton in previous papers*. On the assumption that the frequency of the wave is greater than the frequency of collision ν of the electrons he has shown that

$$\mu^2 = 1 - \frac{2x(1-x)}{2(1-x) - y_T^2 \mp \sqrt{\{y_T^4 + 4y_L^2(1-x)^2\}}}$$

and

$$k = \frac{(1-\mu^2)^2}{2c\mu} \cdot \frac{\nu}{x}\left[1 + \frac{y_T^2}{2(1-x)^2} \mp \frac{y_T^4}{2(1-x)^2\sqrt{\{y_T^4 + 4y_L^2(1-x)^2\}}}\right],$$

x, y_L, y_T where $$x = \frac{4\pi Ne^2}{mp^2}, \quad y_L = \frac{H_L e}{mcp}, \quad y_T = \frac{H_T e}{mcp},$$

H_L, H_T and H_L and H_T are the components of the earth's magnetic field along and perpendicular to the wave normal.

For vertical incidence of 100-metre waves in London we have $y_L = 0{\cdot}406$ and $y_T = 0{\cdot}174$, so that

$$\mu^2 = 1 - \frac{2x(1-x)}{2(1-x) - 0{\cdot}0303 \mp \sqrt{\{0{\cdot}00092 + 0{\cdot}66(1-x)^2\}}}$$

and

$$k = \frac{(1-\mu^2)^2}{2c\mu} \cdot \frac{\nu}{x}\left[1 + \frac{0{\cdot}0303}{2(1-x)^2} \mp \frac{0{\cdot}00092}{2(1-x)^2\sqrt{\{0{\cdot}00092 + 0{\cdot}66(1-x)^2\}}}\right].$$

The transverse component of the applied field exerts little influence on the refractive index and on the absorption coefficient except in the region where $x \doteqdot 1$.

In order to obtain an approximate idea of the diurnal variation of the relative intensity when a magnetic field is operative, the data for the ionization in the E region and for the collisional frequency ν of the electrons given in the previous sections have been used for calculations of the relative intensity for the case of propagation entirely along the field. The intensity of the longitudinal component of the applied field has been taken equal to the vertical component of the earth's magnetic field in London. Thus $y_L = 0{\cdot}406$. For such a case the expressions for the refractive index and for the absorption coefficient are simpler than for the general case. They are

$$\mu^2 = 1 - \frac{x}{(1 \mp y_L)}$$

and

$$k = \frac{(1-\mu^2)}{2c\mu} \cdot \frac{\nu}{(1 \mp y_L)}.$$

The lower sign in these equations refers to the ordinary ray and the upper to the extraordinary ray.

Before the waves are returned from the E region they must pass twice through this region when returned from the F region. The absorption suffered in the E region is then given by

$$2\int k\,ds = \frac{1}{c(1 \mp y_L)} \int \frac{\nu(1-\mu^2)}{\mu}\,ds,$$

* *J.Inst.E.E.* **71**, 642 (1932).

integrated through the total thickness of region E. When they are returned from the E region the waves penetrate a distance y_0 into the region to the point where $\mu^2 = 0$ and return from this point. The total absorption suffered is then

$$= \frac{1}{c(1 \mp y_L)} \int_0^{y_0} \nu \frac{1 - \mu^2}{\mu} dy.$$

For a parabolic distribution of the ionization

$$N = N_0 y^2/y_0^2 \text{ with } N_0 = \frac{mp^2}{4\pi e^2}(1 \pm y_L).$$

This becomes

$$\frac{y_0}{c(1 \mp y_L)} \int_1^0 \nu (1 - \mu^2)^{\frac{3}{2}} d\mu.$$

It should be noticed that the ordinary ray travels a greater distance y_0 in the region than does the extraordinary ray.

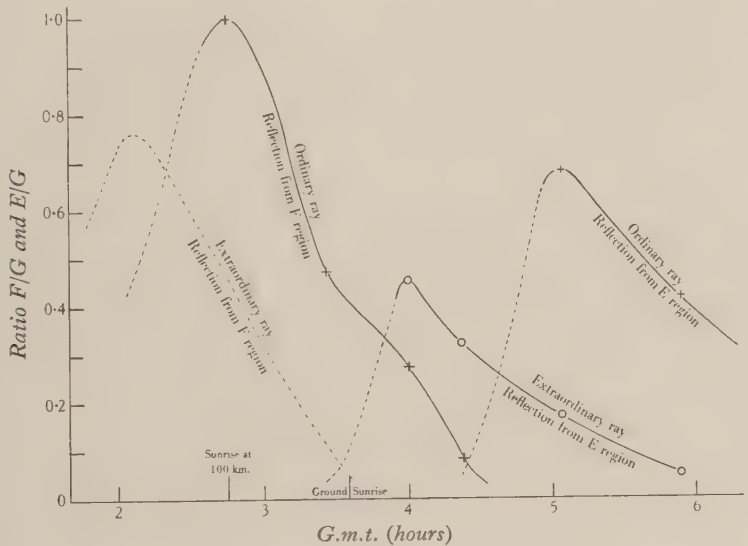


Figure 4. ○ Extraordinary ray. + Ordinary ray.

Assuming that at 2.45 a.m. the ordinary ray is returned from the F region and suffers very little absorption, so that the ratio F (ordinary)/G may be put equal to unity at this time, the diurnal variation of the relative intensity shown in figure 4 is found as a result of these calculations.

With the applied magnetic field operative, the additional features to be looked for in the diurnal variation of the relative intensity are as follows. Before the E region becomes sufficiently ionized, two components of the waves returned from the F region will be received at the ground. With the beginning of ionization in the E region both components will be absorbed, the extraordinary ray suffering the greater absorption. Since fewer electrons are required to cause the return of the extraordinary ray, this component will be first returned from the E region, the ordinary ray being returned at a later time. When either component is first returned

the electron-limitation effect will cause it to have small intensity, but as time advances each component will exhibit a maximum of intensity owing to the decreasing effect of electron-limitation together with increasing absorption in the lower part of the E region. In spite of the fact that the ordinary ray, when reflected from the E region, travels a greater distance in the ionized medium, the absorption for this ray is less than that for the extraordinary ray.

In Chapman's calculations of the ionization it is assumed that a monochromatic radiation from the sun is absorbed in an atmosphere in which the density varies exponentially with height. If the number of ions produced per second is I it is also assumed that the ionization condition is of the form

I

$$dn/dt = I - \alpha n^2,$$

n n being the total number of ions per cm^3 at any time. The recombination coefficient α is taken to be constant for all heights. From these considerations it appears that as a result of the ionization during one day a considerable amount of ionization will remain, after sunset and throughout the night, at the lowest levels reached by the ionization at noon. In figure 1, for instance, the curve for 4 a.m. should not according to Chapman show zero ionization at about 108 km. but should show a small but appreciable amount down to about 85 km. In the above work this residual ionization has been arbitrarily neglected by terminating the curves at a greater height. If the residual ionization at low levels does exist it will introduce, throughout the night, a considerable amount of absorption into the path of the waves reflected from F region. This is contrary to experimental observations, for the intensity of the F reflections is always high during the night, except at those times when it is obvious that electron-limitation in this region is effective.

§ 4. EXPERIMENTAL INVESTIGATION

In the series of experiments presented below, observations have been made of the relative intensity of the reflected wave with respect to the ground wave over a period extending from midnight until some time after sunrise at the ground. Wave-lengths from 100 to 170 metres have been used. In the first experiments the frequency-change method of observing the effective height has been employed. The average value of the relative signal-strength may be arrived at from the photographic records obtained, either by observing the fading or from the fringes that result when the frequency is changed. Fading may occur owing either to actual intensity-variations of the downcoming wave or to interference between the downcoming wave and the ground wave. The short-period fading that causes the signal to vary between maximum and minimum many times in the course of a minute is probably of the interference type, the maximum occurring when the downcoming wave and the ground wave are in phase and the minimum when they are in antiphase. For the former the total signal may be represented by the sum of the intensity F of the downcoming wave and that G of the ground wave, while in the x latter case we have $(F - G)$. If the observed ratio $(F - G)/(F + G) = x$, then $F/G = (x + 1)/(1 - x)$. When a frequency-change is made we purposely bring the

two waves into and out of phase so that the same ratio may be measured. In these experiments the amplifier was calibrated by means of a local oscillator.

The frequency-change method suffers from the disadvantage that when there is more than one downcoming wave the records are complex and it is difficult to sort out the various components. For this reason the Breit and Tuve method, as developed by Appleton and Builder, has been employed in the later runs. With this method the various components are received separately and their relative intensity may be easily measured. It is now well known that the magneto-ionic components of waves of the order of 100 metres return to the earth circularly polarized owing to the influence of the earth's magnetic field on the motion of the electrons in the ionized regions, the extraordinary ray having a right-handed sense of rotation and the ordinary component a left-handed sense, as looked at along the ray in the direction in which the wave is travelling. When the frequency-change method was used and the signals were received on an aerial or on a loop no account was taken of this fact, so that for these results it is the average maximum value of the two components together that has been measured. With the echo method, signals being received on an ordinary loop aerial, the same thing is measured when the two components are not resolved by the layer and the apparatus, but when they are so resolved the average maximum value of each component may be found separately. If, however, a circularly polarized aerial, which will receive the left-handed circularly polarized component to the exclusion of the right-handed component or vice versa, is used, then the average maximum value of each component may always be found if the waves are truly circularly polarized. I am greatly indebted to Messrs Ratcliffe and E. L. C. White for advanced details of this aerial, which has later been described in a paper published in the *Philosophical Magazine**.

§ 5. EXPERIMENTAL RESULTS

Several experimental runs are described and discussed below.

February 12, 1932. Wave-length 170 metres. The frequency-change method was used in this run, the relative intensity being deduced both from the fringes and from the fading with good agreement. The results are shown in figure 5. The

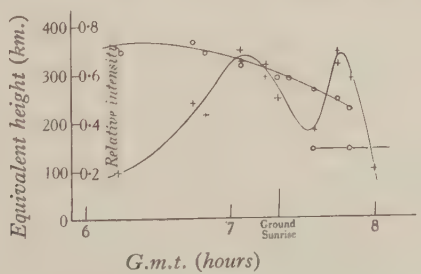


Figure 5. Wave-length 170 metres, February 12, 1932.
+ Relative intensity. ○ Equivalent height.

* *Phil. Mag.* **16**, 125 (1933).

wave from F region was observed just after 6 a.m., its amplitude being then small. Between 6 a.m. and 7 a.m. the amplitude steadily increased. Since it is improbable that an hour would be required for the electron-density to build up to the fully reflecting value in the F region, we may account for some, at least, of this increase by the continued removal of absorption in region E. After 7.05 a.m. a sharp decrease in the intensity was observed, this being due to the increase of ionization and therefore absorption in region E. This decrease continued until 7.35 a.m., after which time another maximum of the relative intensity was observed. This coincides with the transition from F to E reflection. Twenty minutes later the intensity of the wave reflected from region E had decreased to a small value owing to absorption in the lower part of this region. This set of results is typical of the results obtained with the frequency-change method.

June 22, 1932. Wave-length 100 metres. In this run the echo method was used, the signals being received on a loop aerial. In figure 6 (*a*) the relative intensities

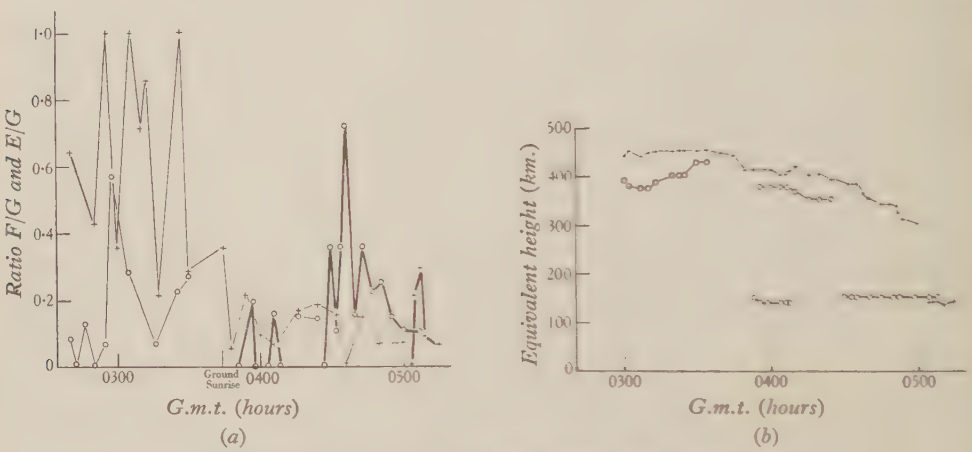
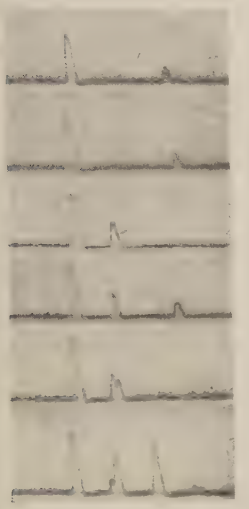


Figure 6. June 22, 1932. Wave-length 100 metres*.

* In this and in the following figures the cross is used for the ordinary ray and the circle for the extraordinary ray. The thin line is used for reflections from F region and the thick line for reflections from E region.

are shown and in figure 6 (*b*) the corresponding equivalent heights of reflection. Before 3.30 a.m. the waves reflected from the F region show a considerable amount of intensity-fading. Reflection from the F region had persisted throughout the night, the ordinary ray being stronger than the extraordinary ray. Just before ground sunrise the intensity of both components reflected from region F begins to decrease owing to the beginning of absorption in region E, and this continues until after the appearance of reflections from region E itself. About 4 a.m. a weak reflection is obtained from region E, but this echo does not give maximum intensity until 4.40 a.m. This is the extraordinary ray for region E, its appearance coinciding with the disappearance of this component from the F region; there is actually a disappearance of the echo between 4.10 a.m. and 4.30 a.m. After the maximum

there follows a decrease in the intensity which continues until 5.10 a.m. At 5.05 a.m. a doublet echo is obtained from the *E* region, figure 7, the more retarded component showing a small maximum (the ordinary ray). After this time the records are



4.15 a.m. Reflection from region *F* showing the extraordinary and ordinary components.

4.25 a.m. Reflection from region *F* showing the ordinary ray only.

4.30 a.m. Reflection from region *E* showing the extraordinary ray.

4.32 a.m. Reflection from both regions showing the extraordinary ray from region *E* and the ordinary ray from region *F*.

5.05 a.m. Reflection from region *E* showing both extraordinary and ordinary components.

5.07 a.m. Reflection from region *E* showing both components with a strong second reflection of the ordinary ray.

→
Time

Figure 7. June 22, 1932.

obscured by interference, but there is no doubt that this latter component is the ordinary ray for the *E* region.

June 17, 1932. Wave-length 98 metres. Reception on a loop aerial. This set of results, figures 8 (*a*) and 8 (*b*), will serve to amplify those of the previous run. During the night, from 1 a.m. until 3.20 a.m., the ordinary ray from the *F* region appeared only occasionally. This is most certainly due to the fact that while the *F* region was a good reflector for the extraordinary ray, the electron-limitation effect caused it to be a very poor reflector for the ordinary ray. At 3.30 a.m. conditions improved for the ordinary ray, but since both components of the doublet were almost superimposed it is not possible to estimate their relative intensity separately. The difference in the equivalent height can, however, be approximately obtained. The total intensity decreased steadily until 4.15 a.m., at which time we observe the first reflection from the *E* region, the extraordinary ray. At 5 a.m. the ordinary ray from the *E* region is observed. In this set of results the maximum for both components is seen*.

* In the two runs described immediately above, the polarization of the downcoming waves was not observed. For the *F*-region reflections there is no doubt about the identity of the two reflected components, the ordinary and the extraordinary rays. The identity of these two components in the *E*-region reflections was for these two runs worked out from the variation of the intensity with time of day and from the fact that a doublet echo was obtained for this region. It seemed desirable to make certain of these facts by using a polarized aerial with which the polarizations of the two components from the *E* region could be found. Such a run has been carried out with a wave-length of 100 metres. Precisely similar results to those described above were obtained.

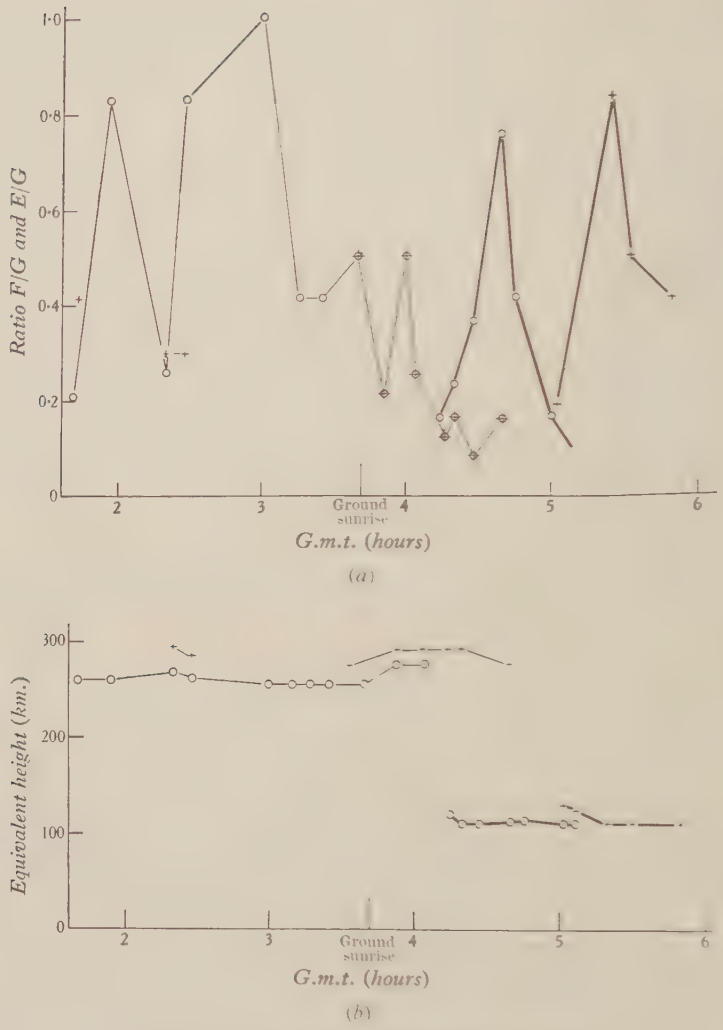


Figure 8. June 17, 1932. Wave-length 98 metres.

March 25, 1933. Wave-length 100 metres. Reception on a circularly polarized aerial. These results, figures 9 (*a*) and 9 (*b*), show very clearly the electron-limitation effect in the *F* region and the ultimate absorption of the waves by region *E*. Transition from *F* to *E* reflection was observed, but on this day the ionization in region *E* was never sufficiently great for this region to return a strong signal. The first reflection from *F* occurred at about the time of ground sunrise; see figure 9. At 5·54 a.m. a signal of small amplitude, the extraordinary ray, was received. The amplitude increased to a maximum by 6.13 a.m. owing to the removal of the electron-limitation, and a consequent increase in the reflection coefficient for this ray. Since this occurred after ground sunrise, ionization in *E* had already started,

so that the subsequent decrease in the intensity of the extraordinary ray must be accounted for by absorption in region *E*. When, at 6.20 a.m., the *F* region begins to reflect the ordinary ray, we observe an increase in its amplitude, in spite of the presence of absorption in *E* region, as the layer becomes a more and more efficient reflector. The absorption of this ray ultimately follows between 6.46 and 7.20 a.m.

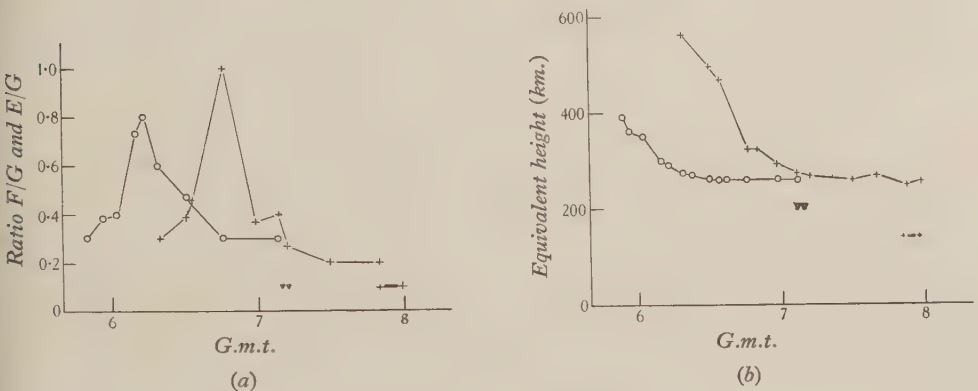


Figure 9. March 25, 1933. Wave-length 100 metres.

At 7.07 a.m. the two components are reflected at very nearly the same height, and at later times the extraordinary ray is absent, presumably completely absorbed in region *E*. At 7.50 a.m. a weak signal was observed from the *E* region, and by that time the relative intensity of the ordinary ray from *F* had been reduced to 0·2. During the period from 6.40 to 7.10 a.m. when the *F* region was a good reflector for both components, the ordinary ray was less absorbed than the extraordinary ray.

§6. DISCUSSION OF RESULTS

In the light of the theoretical and practical results we may summarize the general features of the diurnal variation of the relative intensity.

(1) During the night, when the ionization in *E* is small and decreasing, we have two conditions for short waves. (*a*) *With a wave-length that does not penetrate F region.* A strong signal which steadily increases as the *E*-region absorption is removed will be received, the ordinary ray being the stronger if there is no electron-limitation effect in the *F* region. A wave-length may be found for which the *F* region is a good reflector for the extraordinary ray but, owing to the fact that more electrons are required for the ordinary ray, this ray may suffer electron-limitation. In such a case the extraordinary ray will be the stronger. (*b*) *With a wave-length that does penetrate F region.* As the ionization in region *F* increases with the incidence of sunlight on the layer, the extraordinary ray is first reflected with a small intensity. Its intensity immediately begins to increase owing to the removal of electron-limitation for this ray. At a later time the ordinary ray is reflected and its intensity also increases for a similar reason.

(2) As the ionization in region *E* builds up, the first effect of the renewed

ionization in this region is to cause increased absorption of the waves which, being reflected from region F, pass twice through region E. Under these conditions the extraordinary ray is then absorbed more than the ordinary ray.

(3) When the ionization in the E region has increased sufficiently this region begins to reflect. The extraordinary ray is first reflected. It is found that the ordinary ray is not reflected until about an hour later. The intensity of each of these components is limited by electron-limitation when it first appears and by absorption at a later time, the absorption taking place in the lower part of the E region itself.

§ 7. THE DOUBLET ECHO FROM REGION E

In 1931 Appleton and Builder* pointed out that the doublet echo that they had obtained when a single impulse was reflected by the F region could be accounted for on the magnetic-ionic theory if the equivalent height as measured by the ordinary ray was different from that measured by the extraordinary ray. This explanation of the phenomenon, which occurs very frequently in the case of reflection from the F region, has received ample confirmation since that time. It has since been shown by Ratcliffe and E. L. G. White† that both the extraordinary and the ordinary ray are distinguishable by means of a suitably polarized aerial in reflections from the E region at certain times of day, but that the equivalent height of this region as measured by either component is the same.

On several occasions during the course of the work described above, doubling of the echo from the E region was observed. The separation of the components was always small. These observations have since been confirmed by some recent work of Appleton‡. It has been found that the intensity of the extraordinary ray from the E region is comparable with that of the ordinary ray for only a very short period during the morning, for wave-lengths of about 100 metres, and that this component is probably absorbed during the greater part of the day. This may in some part account for the fact that the doubling of the echo from the E region is very infrequently observed.

§ 8. ACKNOWLEDGMENT

I wish to express my thanks to Prof. E. V. Appleton for his encouragement and help in this work.

* *Proc. Phys. Soc.* **44**, 76 (1932).
† *Phil. Mag.* **16**, 125 (1933).
‡ *Proc. Phys. Soc.* **45**, 673 (1933).

551.51.053.5:621.396.11

SOME AUTOMATIC RECORDS OF WIRELESS WAVES REFLECTED FROM THE IONOSPHERE

By J. A. RATCLIFFE, M.A.

AND

E. L. C. WHITE, B.A., Cavendish Laboratory, Cambridge.

Communicated by Prof. E. V. Appleton, F.R.S., September 25, 1933.
Read November 3, 1933.

ABSTRACT. Some automatic records of wireless waves reflected from the ionosphere are discussed. The records extend over a period of fourteen months. They indicate that the region below the *F* region may be triply stratified and consist of the intermediate region (effective height approximately 120 to 180 km.), the *E* region (effective height approximately 100 to 120 km.) and the *e* region with an effective height which is always nearly 105 km., within $\pm$ 5 km. The *e* region is shown to be intermittent in its occurrence, and is identified with the "nocturnal *E* region" mentioned in previous papers. The temporal variation of the ionization in these three regions is discussed.

A statistical investigation is made to see whether the occurrence of the nocturnal ionization in region *e* is associated with the occurrence of magnetic disturbances or of thunderstorms. It is shown that if the nocturnal ionization was unrelated to magnetic disturbances the probability of the observed coincidences between the occurrence of the two phenomena would be 0·01, and if thunderstorms and the nocturnal ionization were unrelated the corresponding probability would be 0·0015.

§ 1. INTRODUCTION

In a recent paper[1] we have described an apparatus for continuous automatic recording of the equivalent height of wireless echoes returned from the ionosphere, and have given a preliminary account of some of the results obtained. This apparatus has now been in use for fourteen months, and the present paper gives a fuller account of various conclusions we have been able to draw from the data obtained during this period. We have used the apparatus to make weekly records on the international days of the polar year for the Radio Research Board, and have therefore worked, very largely, on the polar-year wave-lengths of 75 and 150 m.

§ 2. FINE STRUCTURE OF THE IONOSPHERE

It was mentioned in the previous paper that, on a wave-length of 100 m. in the summer months, the normal behaviour is for the *F*-region reflection to give place to the *E*-region reflection discontinuously near sunrise; but that later in the year, with a wave-length of 175 m., cases had been observed in which the transition from the *F* to the *E* reflection took place gradually. Further records have shown that this type of behaviour is quite common, and have led to the recognition of

more than two ionized regions in the ionosphere*. The record reproduced in figure 1 illustrates the phenomenon. This record was taken on November 27, 1932 with a wave-length of 150 m. From 0700 to 0840 an *F* echo was recorded, with equivalent height decreasing from 300 to 200 km. From 0840 to 0940 reflection took place from an intermediate region with variable equivalent height of the order of 180 km., and at 0940 reflection from the *E* region at 130 km. was established. At 1410 on the following evening the intermediate region (150 km.) appeared and at 1510 it gave place to the *F* region (220 km.). The record of figure 2 was taken on the previous evening, November 26, on the same wave-length; it shows a similar behaviour, and, in particular, it shows a great equivalent height for the *F* echo just as it comes in. This is a characteristic behaviour of the *F* echo when the intermediate region is present, and may often lead to the detection of the intermediate region on a record; it is presumably due to group retardation in the intermediate region. The presence of the echo at 1410–1430 will be discussed later. The intermediate region has only been observed, on these wave-lengths, near sunrise and sunset; in the middle of the day the ionization of the *E* region is sufficiently intense to prevent the waves from reaching the intermediate region.

During the winter months the equivalent height of the *E* region on a wave-length of 150 m. is somewhat variable, but lies round about 125–140 km. On a wave-length of 200 m. it is lower, at about 105 km. The high value and the variability on the 150 m. wave-length suggest that this is near the critical wave-length during the winter day, in agreement with the critical-wave-length curve of Appleton and Naismith[2]. In the summer months the equivalent height of the *E* region (λ – 150 m.) is less, and of the order of 105 km., and is less variable during the day; moreover the same equivalent height is recorded for a wave-length of 75 m. This seems to show that 150 m. is not near the critical wave-length in the summer.

In the previous paper attention was drawn to the occurrence of nocturnal increases in the ionization of the *E* region. It is now found that the equivalent height of this nocturnal *E* region is remarkably constant from night to night, being about 105 km., whichever wave-length (150 m. or 75 m.) is used. It is found, during the winter months, that daytime echoes may occur sporadically for an hour or two, at any time of day, at the same equivalent height of 105 km. It is very probable that these are due to the type of intermittent ionizing agency which produces the nocturnal *E* region. The echoes in figure 2 between 1410 and 1430 and after 1600 are at an equivalent height of 105 km. (compare the *E*-region height of 130 km. on the same record) and are of this type. Other examples are shown in figure 3† (January 29, 1933) and in figure 4 (December 22, 1932). In the latter figure the 105 km. nocturnal region is present during the sunrise period.

* The presence of more than two ionized regions has also been suggested by Schafer and Goodall[11] and Appleton[12]. Our present picture of the detailed structure of the ionosphere corresponds very closely with that suggested by the latter author at a recent discussion before the Royal Society.

† In making the records of figures 3, 7 and 8 a circularly polarized receiver was employed which was automatically switched so as to receive right- and left-handed polarized waves alternately[6]. A black line at the top of the record corresponds to reception of right-handed polarization.

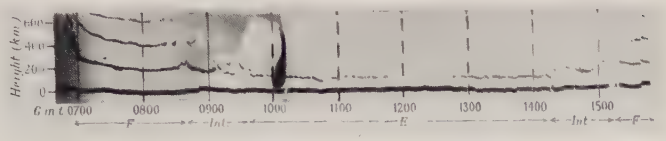


Figure 1. November 27, 1932 (150 m.).

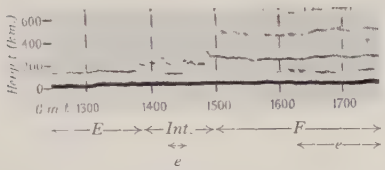


Figure 2. November 26, 1932 (150 m.).

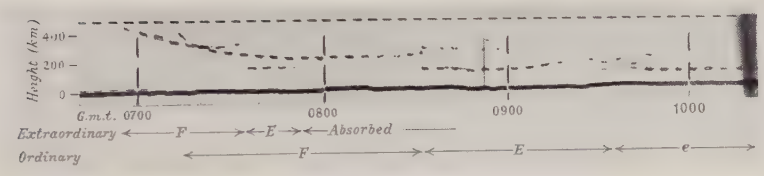


Figure 3. January 29, 1933.

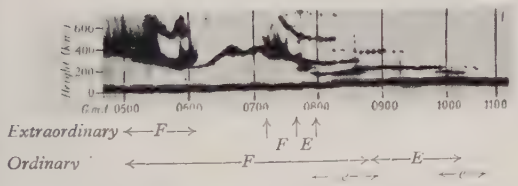


Figure 4. December 22, 1932 (150 m.).

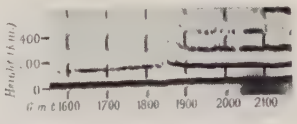


Figure 5. July 7, 1933 (150 m.).

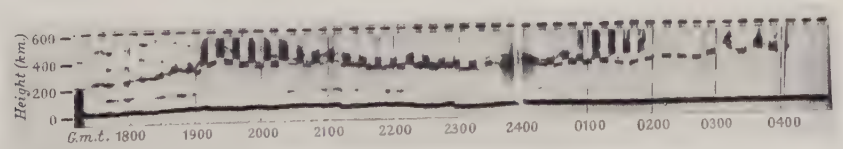


Figure 7. January 26–27, 1932 (150 m.).

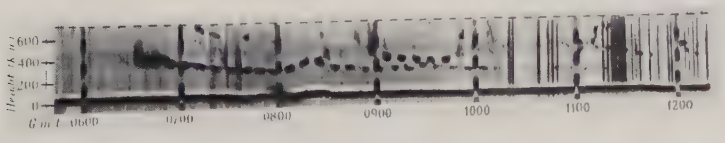


Figure 8. March 22, 1933 (75 m.).

To face page 108

During the summer months the *E* region itself appears, on these wave-lengths of 150 and 75 m., at an equivalent height of about 105 km. and, in consequence, the appearance of extra ionization at this level during the daytime owing to the action of the intermittent ionizing agency passes unnoticed. It is occasionally found, however, that a break occurs in the record of the 105 km. echo, of such a nature as to suggest that the reflection has changed from one level of ionization to another extremely near it. Thus at 1830 in figure 5 (July 7, 1933) there is a break in the record, and the echo then seems to descend rapidly from a greater height to the same level as previously. The phenomenon is repeated at 1845; the faint second-order reflection shows the splitting more clearly. This is very suggestive of a jump from a lower to a higher region, accompanied by a group-retardation effect and a magneto-ionic splitting. The same phenomenon occurred at the same time on eight out of nine records made between July 7 and 18, and on the remaining record (July 10) the *E*-region echo gave place to the *F*-region echo in the usual way (possibly with a little intermediate region in evidence) at this same time, 1830. This seems to confirm the idea that in the summer the true *E* region and the intermittent region are practically coincident.

It seems very probable that the nocturnal *E* region and the intermittent daytime region at 105 km. are due to the same ionizing agency, for they both occur at the same level and are both intermittent in character; it is therefore convenient, following Appleton[4], to give both these the name of *region e*. The possibility of the ionization of region *e* being due to charged corpuscles from the sun, or to thunder-clouds, is discussed in § 4.

In a preliminary account of these results which was contributed by one of the present authors to a discussion at the Royal Society[7], and in a letter to *Nature*[13] in which figure 3 was reproduced, the region at a level of 125–140 km. in winter ($\lambda = 150$ m.) was wrongly interpreted as being the intermediate region. The study of records of the type reproduced in figures 1 and 2 has now shown that this is the *E* region and that the intermediate region is above it and region *e* below it, so that there may be a triple stratification of the region between effective heights of 100 km. and 200 km.

The multiplicity of echoes which is possible from the *E* region, the intermediate region, and the intermittent region *e*, makes it very necessary to be cautious in interpreting split echoes from these levels as due to magneto-ionic effects.

A record obtained on December 14–15, 1932 throws some interesting light on the relation between region *E* and region *F*. During the night of December 14 the *F* ionization was very weak, as is evidenced by the fact that, with a wave-length of 150 m., the extraordinary component of the echo was absent, owing to electron-limitation, for five hours, whereas on other days of this month it was never absent during the night. On the following day reflections from the *F* region alone were visible, the *E* region and the intermediate region being too weak to give rise to reflection. It therefore appears that weak ionization in the *F* region is associated with weak ionization in both the *E* region and the intermediate region, and it is reasonable to assume that the ionizations of these three regions are due to a common

cause. This agrees with Appleton's suggestion that all are due to the effect of ultra-violet light.

§ 3. BEHAVIOUR OF THE *F* REGION

By observing the time at which the waves cease to be reflected from the *F* region we know the time at which the ionization-density has decreased to a certain fixed value. These times are plotted throughout the year for a wave-length of 75 m., for the ordinary and the extraordinary wave separately, in figure 6; the times at which the *F* region begins to reflect in the morning are shown also. On this same figure the times of sunset and sunrise both at the ground and at a height of 200 km. are plotted. In computing the times for 200 km., allowance is made for the fact that the obliquity of the sun's rays varies throughout the year*. It is seen that in summer the ordinary wave begins to be reflected about $1\frac{1}{2}$ or 2 hours after ground sunrise, whereas in winter reflection begins about $\frac{1}{2}$ hour after ground sunrise.

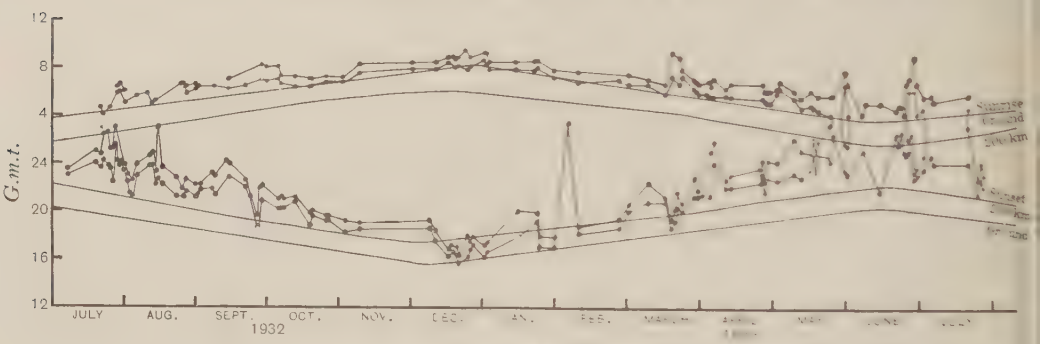


Figure 6.

The effect is noticeable on a similar curve published by Elias, von Lindern and de Vries[8]; they do not, however, comment on it. The phenomenon is very unexpected; we should anticipate just the opposite in view of the fact that the amount of ionization remaining through the night must be considerably greater in summer than in winter. A possible, though improbable, explanation is that the earth is nearer the sun at midwinter than at midsummer, so that the intensity of the sun's light is greater by about 6 per cent.

The records on a wave-length of 150 m., during the winter months when this wave-length is nearly critical, reveal a marked tendency for *F* ionization to increase in density round about midnight. A similar effect has been noticed by Kenrick[9]. In figure 7 (January 26–27, 1933), which is typical of many records, the splitting of the echo between 1900 and 2300 indicates weakness of the ionization, and the rejoining of the split echo near 2300 indicates a reconcentration. The splitting and disappearance of the echoes just before sunrise corresponds to the ordinary decrease in ionization-density at this time. Records on a wave-length of 75 m. in the summer indicate the same effect, but the conclusions to be drawn from them

* We are much indebted to Mr R. H. D. Mayall for help with this calculation.

are not so certain because at a height of 200 km. the night is very short at this time of the year.

Rukop[10] has drawn attention to a phenomenon which he calls the *Abendkonzentration*. He observes that, with sufficiently short wave-lengths, the *F* echo is doubled during the day, but after sunset the two components join together to give a single echo which later splits and finally ceases to be reflected in the usual manner. We, also, have observed this behaviour on several occasions throughout the year, whenever the wave-length used was nearly enough critical for the *F* region. By using the circularly polarized receiver we find that the upper echo during the day corresponds to the "ordinary" left-handed polarized wave; observations of the intensity confirm this, since this echo is always the stronger. It is possible that this daytime splitting is due to group retardation in the ledge which Appleton[4] has shown to exist on the lower side of the *F* region, and that the disappearance of the splitting in the evening is due to the disappearance of this ledge by recombination. Appleton's figure 2 shows the same kind of midday splitting on a wave-length of 75 m. A similar phenomenon, often observable at the sunrise period, is illustrated in figure 8 (March 22, 1933) and can probably be explained in the same way. At 0630 the extraordinary echo appears and at 0850 the ordinary, which descends nearly to the level of the extraordinary at 0930 only to rise again immediately and stay at a considerably greater height than the ordinary until 1220, when the record finishes. This record, taken with a circularly polarized receiver, illustrates the nature of the polarization of the two echoes. If the present explanation is the correct one, there is no need to postulate the presence of an evening concentration of ionic density as suggested by Rukop.

§ 4. NOCTURNAL *E* IONIZATION

In the previous paper it was suggested that nocturnal increases of the ionization in region *e* might be due to charged solar corpuscles, such as are supposed to produce disturbances in the earth's magnetic field; or to the action of thunder-clouds, as suggested by C. T. R. Wilson. We have now sufficient data to investigate whether there are any correlations between the nocturnal increases of ionization and magnetic disturbances or storm clouds. Appleton and Naismith[2] have pointed out that they have very often found magnetic activity to be associated with an increased ionization "in the Kennelly-Heaviside layer," and other workers have occasionally noticed the same phenomenon[3]. A preliminary examination of our records shows, however, that there is by no means a one-to-one relation between the occurrence of nocturnal *e* ionization and either thunderstorms or magnetic disturbances, and that a statistical investigation is necessary before any relation can be established. Thus on the night (April 30–May 1, 1933) of commencement of the largest magnetic disturbance in the period under review, there was no record of nocturnal ionization in region *e*, even on a wave-length of 150 m. which was not far from being critical for the *E* region. On several occasions records on 75 m. wave-length showed absence of nocturnal region *e* ionization even when world-wide magnetic storms were recorded; since, however, 75 m. is far from being the

critical wave-length for the daytime region, these observations are not particularly significant. On the other hand, nocturnal-region-*e* ionization was often recorded, on both wave-lengths, when the night was magnetically quiet. In view of these results it seems essential to make a statistical analysis before any relation can be established.

In searching for a relation between the nocturnal ionization in region *e* and magnetic disturbances, we have distinguished between magnetically quiet and magnetically disturbed nights by taking the range of the variation of the magnetic declination through the hours of darkness and comparing it with the mean value of this range over the month in which the night occurred. Nights with a range greater than the monthly mean were classed as disturbed, and those with a range less than the mean as quiet. The magnetic data were taken from the figures published by the Abinger Magnetic Station, Surrey.

In the present experiments the wave-lengths used were 75, 150 and 200 m. It was found that the wave-lengths 150 and 200 m. nearly always showed some echo from region *e* or *E* during some part of the night; nocturnal-region-*e* ionization was only considered to be present if the ionization was sufficiently strong to prevent the signal from reaching the *F* region, or if the *e* echo was recorded during a period when at least one of the magneto-ionic components of the *F* echo was absent on account of electron-limitation. The nocturnal-*e* ionization was considered to be absent if the total duration of the *e* reflection, between the time when the daytime *E* echo disappeared in the evening and the time when it reappeared in the morning, was less than 15 minutes. This classification excludes several of the records on 150 and 200 m. which show simultaneous echoes from the *e* or *E* and *F* regions. On the wave-length 75 m. the borderline cases are much less numerous, and the classification into "nocturnal *e* present" and "nocturnal *e* absent" is simple. Table 1 summarizes the results which have been found. The numbers represent frequencies of occurrence, e.g. the number in the left-hand top corner is the

Table 1.

	Magnetically disturbed	Magnetically quiet
Nocturnal *e* present	39	55
Nocturnal *e* absent	19	63

$\chi^2 = 6{\cdot}7$, $p = 0{\cdot}01$.

number of nights which were disturbed and also showed nocturnal-*e* ionization. We find that of the disturbed nights 67 per cent show nocturnal-*e* ionization, whereas of the quiet nights only 46·5 per cent show it. This indicates that if the night is disturbed there is a greater tendency for nocturnal-*e* ionization to be recorded. The fact that so many quiet nights show nocturnal ionization is perhaps to be expected in view of the fact that the occurrence of the nocturnal-*e* echo merely indicates an increase of the ionization density in a somewhat localized region

immediately above the point of observation, whereas an increase of ionization would have to take place over a fairly large region before it could produce an appreciable effect on the earth's magnetic field. If, in spite of this, the results show a relation between the two phenomena, then we may conclude with some certainty that the magnetic disturbance is related to the nocturnal-*e* ionization.

In order to decide whether table 1 indicates any relation between the occurrence of nocturnal-*e* ionization and magnetic disturbances we use the statistical χ^2 test of independence*. The table gives a value of χ^2 equal to 6·7 corresponding to a probability of 0·01 that the same distribution should occur by chance if the two observed effects were completely independent. The conclusion to be drawn is that magnetic disturbances are very probably related to increases of ionization in the *e* region. Although the observations here described are concerned only with nocturnal increases of ionization, it is not suggested that daytime increases are absent.

Observations of the kind here described cannot be used to determine the cause of the ionization, but it is interesting to note that there is a marked tendency for the ionization to occur at nearly the same times on successive nights; this phenomenon would favour the theory of a corpuscular jet radiating from the sun.

It is next of interest to examine whether any relation is found between the occurrence of thunderstorms and of the nocturnal-*e* ionization, as might be expected from Wilson's (5) theory. The results of the year's observations are summarized in table 2, in which the occurrence of thunderstorms in the British Isles is compared with the occurrence of the nocturnal ionization.

Table 2.

	Thunder	No thunder
Nocturnal *e* present	34	60
Nocturnal *e* absent	12	70

$\chi^2 = 10$, $p = 0{\cdot}0015$.

The table shows that 74 per cent of the thunderstorm days showed nocturnal-*e* ionization, whereas ionization occurred on only 46 per cent of the days with no thunder. Applying the χ^2 test we find a value of χ^2 equal to 10 and corresponding to a probability of 0·0015 that this distribution should occur by chance, if thunderstorms and the nocturnal ionization were unrelated.

The fact that the nocturnal-*e* region ionization is correlated with both thunderstorms and magnetic storms suggests immediately that the double correlation may be due to a direct relation between thunderstorms and magnetic storms. The data which have been used in drawing up tables 1 and 2 show no such relation; a similar table comparing the occurrence of thunderstorms with that of magnetic storms gives a value of χ^2 equal to 0·017 and corresponding to a probability of 0·9 that the observed distribution would occur by chance if the two phenomena were unrelated.

* See, for example, R.A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, 1930).

§ 5. ACKNOWLEDGMENTS

In August 1932 the recording apparatus was completely rebuilt, with a more powerful transmitter and a more robust receiving arrangement, with the assistance of the Department of Scientific and Industrial Research, who, on the advice of the Radio Research Board, have made a loan to us of the necessary apparatus. They have also provided us with cathode-ray oscillographs and photographic paper. We wish to take this opportunity of thanking them for their help.

We wish also to express our thanks to Prof. Appleton for many helpful discussions in connexion with the interpretation of the records; to Dr Wishart, of the Cambridge University Department of Agriculture, for help with the statistical analyses; to Prof. Stratton for continuing to provide us with an ideal site for our transmitter at the Solar Physics Observatory; and to Mr J. L. Pawsey for considerable help with the recording apparatus.

REFERENCES

(1) J. A. Ratcliffe and E. L. C. White. *Proc. Phys. Soc.* **45**, 399 (1933).
(2) E. V. Appleton and R. Naismith. *Proc. Phys. Soc.* **45**, 389 (1933).
(3) Hafstead and Tuve. *Terr. Mag.* **34**, 39 (1929).
(4) E. V. Appleton. *Proc. Phys. Soc.* **45**, 673 (1933).
(5) C. T. R. Wilson. *Proc. Phys. Soc.* **37**, 32 D (1925).
(6) J. A. Ratcliffe and E. L. C. White. *Phil. Mag.* **16**, 125 (1933).
(7) Discussion on the Ionosphere. *Proc. R.S.* **141**, 697 (1933).
(8) Elias von Lindern and de Vries. *Tid. v.h. Radiogenootschap*, **6**, 53 (1933).
(9) Kenrick. *Physics*, **4**, 194 (1933).
(10) Rukop. *Electr. Nach. Tech.* **10**, 41 (1933).
(11) J. P. Schafer and W. M. Goodall. *Nature*, **131**, 804 (1933); **132**, 521 (1933).
(12) E. V. Appleton. *Nature*, **131**, 873 (1933).
(13) J. A. Ratcliffe and E. L. C. White. *Nature*, **131**, 873 (1933).

DISCUSSION

Prof. A. O. Rankine. May I seize this opportunity of getting from the lips of experts an explanation of what appears to be an inconsistency in relation to wireless transmission? It is very commonly believed that signal-strength is greater across the night side of the earth than the day side. Yet we have just learned, in this and the previous paper, that reflection from above is exceptional or feeble during the night, owing to insufficient ionization in the appropriate layer. If vertical echoes are stronger during the daytime, how is it that oblique ones are apparently weaker?

Mr A. E. Bennett took exception to the use of the term "echoes." He suggested that the impulses received back from the ionosphere are re-radiations due to resonance.

AUTHORS' reply. Prof. Rankine's question is best answered by a reference to Mr F. W. G. White's paper. Mr White shows that the reflected wave may be unduly weak for two different reasons. Either it may be weak because there is not a sufficient electron-density in the reflecting region to return it to the earth—this is electron-limitation; or it may be weak because the electron-density in the absorbing regions of the ionosphere is so great that, although the wave is reflected, it is strongly absorbed—this is absorption-limitation. Thus, for vertical incidence, a wave may be weak during the day on account of absorption-limitation, and weak or absent at night owing to electron-limitation. For more oblique incidence, however, the electron-density required to produce reflection is not so great, and, for the long-distance transmission referred to by Prof. Rankine, electron-limitation does not usually occur, so that by night the signals are quite strong.

534·14

THE PRODUCTION OF SOUNDS FROM HEATED METALS BY CONTACT WITH ICE AND OTHER SUBSTANCES

By MARY D. WALLER, B.Sc., F.Inst.P.
Lecturer in Physics, London (R.F.H.) School of Medicine for Women.

Received July 1, 1933. *Read, with demonstration, November* 3, 1933.

ABSTRACT. Experiments in which very loud notes were produced as a result of contact of metal bodies with solid carbon dioxide have been described in a previous communication*. The present paper is concerned with the conditions under which soft notes may be produced from heated metal bars brought into contact with ice and a number of other substances. These substances must either sublime or boil or decompose with the evolution of gas, at temperatures for which the metal to be excited still retains adequate vibrating properties. The experiments confirm the theory regarding the mechanism of the phenomenon which was developed in the previous paper. They establish the general conditions under which elastic vibrations of audible frequency may be excited in a metallic body, when contact with another cooler solid substance results in the production of gas from the latter.

§ 1. INTRODUCTION

The conditions under which loud sustained notes may be produced by bringing metal bodies into contact with solid carbon dioxide have been fully described in a previous communication†. This phenomenon, as has already been mentioned in that paper, led to the discovery that small momentary notes may be produced from heated metals when brought into contact with ice, a fact which would otherwise almost certainly have remained unnoticed.

Some of the conclusions reached in the above communication will now be reviewed, and the observations which have been made on several heated metal bars — ice and a number of other materials being used to set them vibrating—will then be described.

(i) Solid carbon dioxide excites metal bars whose natural vibration frequencies may vary from about 1000 to 15000 ~. In preliminary experiments a few of these bars differing widely in frequency were heated, and it was found possible to set them into vibration by contact with ice. There is thus in the case of excitation by ice (or presumably by the other substances used) a wide range of audible frequencies which can be produced.

(ii) Variously shaped bodies -bars, rings, tubes, bells, discs, etc.—can be set into vibration by contact with solid carbon dioxide. It has been considered sufficient to restrict the present study to metal bars, and to assume that bodies of other shapes could be excited if heated suitably.

* *Proc. Phys. Soc.* **45**, 101 (1933). † *Ibid.*

(iii) The phenomenon depends upon the transference of heat from metal to solid carbon dioxide. If metal bars are to be excited by ice, etc. they must be raised to a suitable temperature.

(iv) The body must possess adequate vibrating properties. In the case of excitation by ice or other suitable materials, the variation of this property with temperature has therefore had to be considered; see § 3 below.

(v) The mechanism of the excitation of vibrations, depending upon the sublimation of solid carbon dioxide with the production of considerable gas pressures when it is brought into light contact with the metal, would appear to be somewhat similar to the production of edge tones.

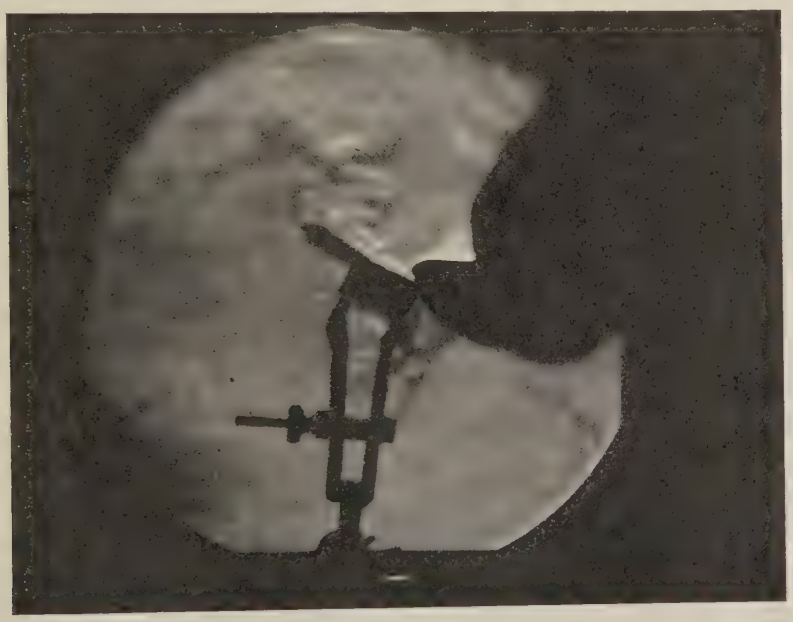


Figure 1. Schlieren photograph of carbon-dioxide sublimation caused by contact with edge of tuning fork.

(By courtesy of the Research Laboratory, General Electric Company, Ltd.)

Through the courtesy of the Research Laboratory of the General Electric Company, Ltd., the author has been able to study the effect on the quantity of gas produced of varying the degrees of contact between a metal bar or tuning fork and a block of solid carbon dioxide, by the Schlieren method, and a photograph very kindly taken by that Company is reproduced in figure 1. The photograph is of a fork viewed end-on, so that the shadows of the prongs are seen in cross section and merge into one. The solid block of carbon dioxide is touching an edge near the tip of one of the prongs, and two streams of carbon dioxide can be seen emerging in definite directions at right angles to the length of the fork. The block was pressing slightly on the edge, and the fork was chattering as well as vibrating regularly. When the contact is light and a pure loud note is being emitted, so little gas is

produced that the two gas streams are barely visible. This fact is of interest and is to be expected, for the contact is slight and intermittent, the condition is one of resonance, and since the damping of the fork is small little energy is required to maintain the vibration. The frequency of the fork being 3000 ~ and photographic exposure $\frac{1}{50}$ sec., sixty contacts take place during exposure, and individual puffs of gas cannot of course be observed.

The Schlieren method of observing the phenomenon thus confirms the theory of the production of the vibrations previously advanced, but shows how little sublimation occurs when loud pure notes are being produced. It would be of interest to observe the Schlieren image through stroboscopic apparatus in order to determine whether the puffs are on alternate sides as they are in the case of edge tones.

§ 2. EXPERIMENTAL ARRANGEMENTS

The bars used were 15 cm. long and of $\frac{1}{2}$ in. square cross-section. They were suspended from two nodes so as to vibrate transversely as free-free bars. The dimensions chosen are convenient for purposes of heating by means of either a blow-pipe flame or an electric oven. Such bars cool at a convenient rate, their vibrations persist for a suitable length of time, and lastly their frequency of vibration is in the region of greatest sensitivity of the ear, which is especially desirable as the notes produced are very soft.

Measurement of temperatures. In preliminary experiments the temperatures were roughly estimated by dropping the bars into a calorimeter containing a known mass of water. It was found for example with a brass bar, that the range of temperature over which notes can be produced by contact with ice was roughly from 100° to 200° C. The notes are accompanied by the hissing of the steam and are quickly quenched by water. At lower temperatures the metal eats its way silently into the ice block, producing water.

In some experiments in which the bars were heated in an electric oven the temperatures were measured by means of a mercury thermometer placed in the oven. It is to be noted that the loudness of the notes produced is not appreciably altered, although the surrounding air as well as the metal is now hot.

Another method for measuring temperatures, used in the case of a brass bar, was to solder a brass and constantan wire at the two nodes respectively, so that the wires fulfilled the double function of suspending the bar and enabling its temperature to be determined thermoelectrically by connecting the two wires with a shunted galvanometer. The wires were soldered into small holes bored into the upper surface of the bar, and interfered but slightly with its vibrating properties.

In most experiments two identical bars, one the vibrator and one the measurer of temperatures, were used. They touched along one edge while being heated, and were subsequently separated so that the former could be conveniently set vibrating. With a bar 15 cm. long it is easy to ensure that the temperature along it is practically uniform. In experiments with a bar, the temperature of which was measured at

the centre and one end by boring two holes and inserting two thermometers in them, the best routine for heating by means of a gas flame was established, and it was found that the temperature of the two thermometers tended to become sensibly equal fairly rapidly even if one was 5° or 10° C. above the other to begin with. The arrangement proved very satisfactory and experimental errors are generally much less than 10 per cent.

Detection of sounds produced. The sounds produced by means of any of the substances tested, other than solid carbon dioxide, are soft and of short duration. They can be heard in a quiet room, and the results given below were detected by ear alone. The author is indebted to Mr G. E. H. Rawlins for constructing a stand for holding the bar and a suitable arrangement for amplifying the sounds, for purposes of demonstration. A light microphone attached to the bar was eventually replaced by a more satisfactory arrangement of ear-phone coils and magnets which were fixed opposite a small piece of iron attached to a side of the bar near one of its ends. The coils were connected to amplifying valves and a loud-speaker.

By this means it would doubtless be possible to extend the list of compounds found to produce notes in heated bars, for the arrangement not only amplifies the note but also eliminates extraneous noises due to boiling, etc., which otherwise tend to mask it.

§ 3. VIBRATING PROPERTIES OF METALS AT DIFFERENT TEMPERATURES

A note on this subject follows the present paper*. The vibrating properties of metals usually decrease with temperature, and the upper limit of temperature for which it is possible to produce notes by contact of a metal bar with ice or other exciting substance is usually determined by this physical factor.

The results obtained for commercially pure copper, aluminium and zinc are shown in figure 2 and for medium steel in figure 3, where the vibrating properties, as determined by the duration in seconds of audibility of a note emitted by the vibrating bar, are plotted against the corresponding temperatures. The bars were struck with a hammer and the time for which the sound remained audible was noted by means of a stop-watch.

It will be noticed that the duration of audibility becomes too small to be measured by means of a stop-watch in the case of the particular samples of zinc, aluminium and copper at about 50°, 150° and 250° C. respectively.

Thereafter the melodiousness of the note decreases with rising temperature until at a higher temperature the pitch is no longer recognizable. It is to be noted that the frequency of the note varies but slightly at different temperatures, since the temperature variations of elasticity and density are both small. The vibrating properties of mild steel reach a minimum at about 120° C. and a maximum at about 240° C., after which there is a gradual decrease with rising temperature, the pitch however being still recognizable at a dull red heat.

* Page 124 of this volume.

§ 4. EXPERIMENTAL RESULTS

The metals used for establishing the conditions necessary for the production of notes in heated bars when brought into contact with various exciting materials were zinc, aluminium, copper, mild steel (or steel), and a number of the results

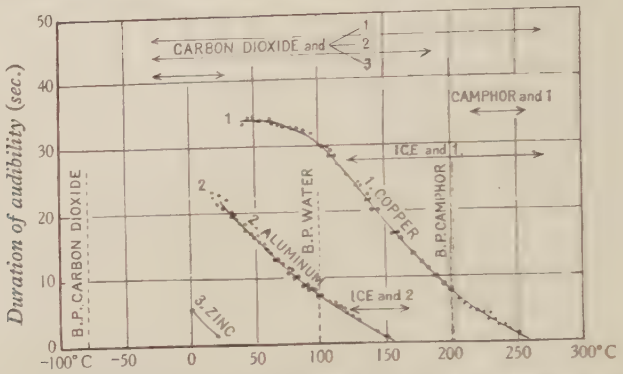


Figure 2. Vibrating properties (as measured by the duration of audibility of notes) of (1) copper, (2) aluminium, (3) zinc at different temperatures. The horizontal lines show the range of temperatures for which notes may be obtained from these metals when brought into contact with various exciting substances.

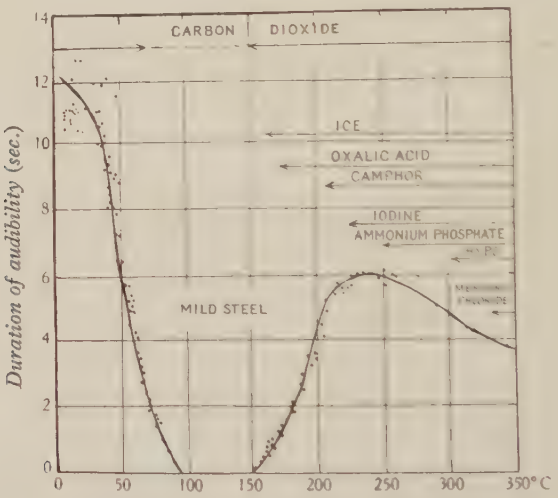


Figure 3. Excitation of notes from heated mild steel by contact with various exciting substances.

obtained are shown graphically in figures 2 and 3. The horizontal lines show the approximate range of bar-temperatures necessary for the production of notes with various pairs of metals and exciting substances. We proceed to consider some of the exciting materials used.

Ice. It is impossible to excite notes from zinc by means of ice, whereas carbon dioxide will produce loud notes. Aluminium, copper and most metals can however

be excited to give soft notes if they are heated above 120° C. and brought into contact with a block of ice. The lower limit of temperature, about 120° C., does not differ very greatly from the temperature necessary to produce the spheroidal state, which is about 140° C.

As is shown graphically in figure 2, the upper limit of temperature is determined by the diminution in the vibrating properties of the metal and is about 150° C. for aluminium and 250° C. for copper. Reference to figure 3 shows that in the case of mild steel the lower limiting temperature, about 160°, depends upon the poor vibrating properties at that temperature and not on the boiling-point of water. The upper limit is not much below the temperature of red heat.

Contact between the ice and metal should be light. The note is somewhat masked by the hissing of the steam (unless it be listened to through the amplifier), but while it lasts it is pure and the phenomenon is similar to that of excitation by solid carbon dioxide, though on a very much smaller scale. The rapid formation of water soon quenches the soft sound.

Camphor, $C_{10}H_{16}O$, passes through a pasty condition in passing from the solid to the gaseous state, its melting-point being 170° C. and boiling-point 204° C. Care must be taken to prevent a flow of hot camphor over the fingers. As is shown in figure 2, notes cannot be produced in zinc or aluminium and only over a very limited range of temperatures, from about 220° C. to 260° C., in copper. Mild steel will be excited between about 220° C. and a much higher temperature.

Iodine, the melting-point of which is 116° C. and boiling-point 184° C., is harder to manipulate on account of the small size of the crystals. Notes have however been produced in mild steel above 230° C. (see figure 3) by holding the crystals by means of forceps.

We have seen that vibrations are produced in a metal bar on account of the gas pressure which results from the sublimation or boiling of the several exciting materials so far considered. It would seem probable, therefore, that materials which decompose with the evolution of gas when brought into contact with heated metals would also be capable of exciting vibrations in heated bars. Again, it might be possible to produce vibrations by means of crystals which give up their water of crystallization when brought into contact with the heated bar.

An exhaustive search on these lines has not been undertaken, but the results obtained with a number of crystalline compounds show that gas evolved as a result of chemical decomposition may produce soft notes over suitable ranges of temperature, whereas the vapour produced as a result of the liberation of water of crystallization is not sufficiently vigorous in its action to produce audible vibrations.

The experiments were generally made on a heated mild steel or steel bar, since these possess good vibrating properties at higher temperatures than the other metals used. The principal results obtained are summarized below and in some cases appear graphically in figure 3.

Oxalic acid, $C_2H_2O_4, 2H_2O$, produced notes when brought into contact with the bar heated to temperatures above about 175° C., on account of sublimation and decomposition with the liberation of carbon-dioxide gas.

Mercuric chloride, $HgCl_2$, produced notes when the bar-temperature exceeded 330° C. Mercuric chloride is generally said to sublime, its melting-point being 277° C. and boiling-point 304° C. Wet drops which evaporate rapidly form on the bar when this is touched on its upper surface.

Ammonium phosphate, $(NH_4)_2.HPO_4$, produced notes when the bar-temperature exceeded 250° C. These are due to decomposition with the liberation of ammonia gas. A boiling deposit which forms on the bar damps the vibrations.

Sodium thiosulphate or "*hypo*," $Na_2S_2O_3.5H_2O$, produced notes, which are probably due to sulphur vapour, above 300° C.

Zinc nitrate, $Zn(NO_3)_2.6H_2O$, produced notes when the bar-temperature exceeded 150° C., probably on account of nitrous fumes and oxygen.

Meta or metaldehyde, which sublimes between 112° and 115° C., will produce notes in mild steel above about 240° C. The block has to be pressed against the heated metal and the note produced, though easily heard at higher temperatures, is not pure. In the case of meta we have passed a long way from the production of notes by solid carbon dioxide, in which (i) contact must be the lightest possible; (ii) the solid carbon dioxide offers a hard background; (iii) the sublimed carbon dioxide remains in the gaseous condition instead of immediately condensing into flakes; (iv) the notes produced are pure, loud and sustained.

Examples of substances which have failed to excite audible vibrations are: benzoic acid, potash alum, manganese sulphate, sodium sulphate, copper sulphate, borax, sodium carbonate, potassium nitrate, ammonium chloride, and ammonium carbonate. The results obtained will not be described in detail; each material behaves individually, and there may be hissing or silence, a deposit due to dehydration or decomposition which sticks on the bar, or a mass which gradually melts. Again if chemical decomposition occurs at a high temperature, for example in the case of potassium bromide, there is no effect, just as a material with an elevated boiling-point would produce no effect. Finally it is necessary that the crystals should be of a reasonable size: notes cannot be produced by means of powders.

§ 5. CONCLUSIONS

(i) Light contact made between two solid bodies which are at different temperatures may result in the production of elastic vibrations of audible frequency in the hotter body.

(ii) The hotter body must be a good thermal conductor, be capable of vibrating at suitable frequencies (about 1000 to 10,000 ~) and possess adequate vibrating properties at temperatures which exceed by some degrees the subliming, boiling or decomposing-temperature of the cooler exciting substance. Vibrations are not produced unless the metal is heated above these temperatures. The upper limit of temperature for which notes may be excited is determined by the loss of vibrating properties of the metal.

(iii) Apart from solid carbon dioxide, which would appear to be unique (at any rate among materials at present available) in its capacity to produce very loud notes,

there are a considerable number of materials which may be used to excite soft pure notes of short duration in heated bars. The mechanism of production is identical with, but on a very much smaller scale than, that of solid carbon dioxide.

Such materials are: (*a*) Those substances which sublime or boil at suitable temperatures when brought into contact with heated metals, for example ice, camphor, iodine, and mercuric chloride. If the sublimation is immediately followed by a return to the solid state in the atmosphere, as in the case of "meta," contact must be heavier and the note produced is not pure. Again the formation of liquid, as in the case of ice, soon quenches the note. (*b*) Substances which decompose when brought into contact with the heated metal with the production of gas pressure, as for example oxalic acid, ammonium phosphate, sodium thiosulphate and zinc nitrate. Solid deposit, or a melting or boiling mass on the metal, is a frequent cause of failure to produce notes with certain other compounds.

(iv) Notes have not been produced when certain compounds containing water of crystallization have been dehydrated by contact with the heated metal.

(v) If a more comprehensive study were contemplated of chemical compounds which will produce notes from heated metals it would be desirable (*a*) to amplify the notes and isolate them from extraneous noises due to boiling, etc., as described in the paper; and (*b*) to use a bar made of a metal which will vibrate at high temperatures, for example steel or monel metal or invar.

DISCUSSION

See p. 126.

534·37

NOTE ON THE VIBRATING PROPERTIES OF METALS AT DIFFERENT TEMPERATURES

By MARY D. WALLER, B.Sc., F.Inst.P.

Received July 1, 1933. *Read November* 3, 1933.

ABSTRACT. A simple acoustical method of studying the vibrating properties of metals at different temperatures is described. Since the damping of the vibrations, which is mainly due to internal friction or solid viscosity, is greatly altered either by previous heat or mechanical treatment and by impurities, i.e. is not constant at any given temperature, such a method capable of giving numerous comparative data over wide ranges of temperature should be of value. Furthermore, there is promise that on account of the large variations of the vibrating properties with temperature it will be possible to obtain, by observations of irregularities in these variations, much interesting information regarding the state of metals and alloys at different temperatures.

The variation with temperature in the vibrating properties of several metals has been studied incidentally in the preceding paper* and a few results are there shown in figures 2 and 3.

The simple hand-and-ear method of observation by which these results have been obtained is now being used to carry out a survey of the vibrating properties of pure metals and alloys of known type between the temperatures of liquid air and red heat.

The author since developing the method has learnt that in 1911 F. Robin† used an essentially similar acoustical method for studying the vibrating properties of various metals between room temperature and red heat, and he was the first to draw attention to what he called the "aphonia" of carbon steels at about 120° C.‡. He used a standard impulse to set the bars vibrating, but the author finds that on account of the logarithmic nature of the damping, the magnitude of the impulse given to the bar may vary over wide limits without appreciably altering the duration of audibility of the note.

The method as used by the author consists in suspending a bar of square or circular cross-section, of side or diameter 0·5 in. and length 15 cm., from two nodes so that it may vibrate transversely as a free-free bar, heating it to known temperatures§, striking it a smart blow with a hammer held in the hand, and determining, by means of a stop-watch, the number of seconds for which the resulting note remains audible to the same observer listening near the bar with the same ear. Properly employed the method is satisfactory, and the degree of accuracy can be gauged by

* Page 116 of this volume.

† Iron and Steel Institute, *Carnegie Scholarship Memoirs*, **3**, 125 (1911).

‡ See figure 3 of the preceding paper.

§ *Ibid.* § 2.

the lie of the experimental points on the curves. The irregularities may prove to be of significance, and not merely due to experimental errors. Losses of energy to the supports, provided these are suitably chosen, or on account of radiation of sound are of secondary importance, but as the research develops they may have to be considered. The most serious damping is due to formation of water drops below the dew-point, and to a lesser extent to the formation of hoar-frost in bars which are cooled to low temperatures in liquid oxygen. If comparisons between different bars are to be made, the frequency of the bar must be taken into account in order to determine the damping per cycle and to allow for varying sensitivity of the ear to different frequencies.

Again, no comparisons or specific constants of different metals can be given unless the exact composition, the degree and length of heating, annealing, cooling, etc., and the amount of working on the metal are known. Thus, for example, commercially pure copper bars purchased from two different dealers, and practically identical as regards dimensions, elasticity and density, vibrated for 3 and 20 seconds respectively at room temperature.

The field that may be investigated by the method would appear to be a wide one. The vibrating properties vary by very large amounts as the temperature changes, whereas many physical properties change but slightly with temperature. If relations can be established between the vibrating properties and certain physical conditions which are hard to detect, the method may be of practical use in the workshop. Properties which depend not only on the composition but on the treatment of the material, such as hardness, tensile strength, fatigue limit, etc., may bear some relation to the vibrating properties.

It is interesting to compare figure 3 of the author's paper cited above with Cuthbertson's* curve for the fatigue limit at different temperatures of medium carbon steel, which shows a very sharp minimum in the fatigue limit at 123° C. Again, the variation of the specific heat with temperature is such that a sharp maximum occurs at about 120° C. in the curve given by Dearden†. The curves given by Thompson and Whitehead‡ for the variations in specific resistance and thermoelectric power include maxima at this same temperature. Sauveur and Lee's§ figure 1 relating tensile strength and temperature of iron and steel shows a marked minimum at 100° C. for electrolytic iron. Herbert in the discussion on Cuthbertson's paper‖ compares the working properties of iron at 120° C. to those of stiff putty or plasticine.

These considerations are sufficient to indicate that a study of the vibrating properties by the method indicated, simple and merely approximate though it may be, is worth while. Its chief virtue lies in the fact that very numerous observations may be made on a property which varies greatly when either the temperature, the composition or the treatment of the material is varied. Irregularities obtained in

* *J. Iron and Steel Inst.* **126**, 237 (1932).
† Iron and Steel Institute, *Carnegie Scholarship Memoirs*, **17**, 89 (1928).
‡ *Proc. R.S.* A **102**, 587 (1923).
§ *J. Iron and Steel Inst.* **112**, 324 (1925). ‖ *Loc. cit.* p. 260.

curves promise to yield much interesting information in view of the present rapidly increasing knowledge of the metallic state.

It may also be possible to compare results with those given by more refined methods for the viscosity of various metals. These are limited in number and often contradictory, see for example the summary given by Erk*. The contradictions are not surprising once it is fully realized that whereas the viscosity of gases or liquids has a definite value for any given temperature, the viscosity of a metal can be enormously altered by mechanical and heat treatment, and by the addition of very small amounts of impurity. Ultimately no study of the damping forces or viscosity of metals will be complete without a knowledge of the crystalline condition of the particular specimen.

DISCUSSION

Prof. Martin Knudsen. I should like to point out the existence of the radiometer force which exists between two bodies in a gas when they are at different temperatures. Thus, suppose that we have two bodies at temperatures θ_1 and θ_2 in a gas at a pressure p, then the radiometer force exerted on every square centimetre of the opposite faces of each is equal to $\frac{1}{2}p\{\Lambda(\theta_1, \theta_2) - 1\}$, provided that their distance apart is negligible compared with the mean free path of the gas.

Suppose now that $\theta_1 = 800°$ K. and $\theta_2 = 200$ K. The radiometer force is then equal to $\frac{1}{2}p$, so that if the pressure be atmospheric a force equal to half an atmosphere is exerted on every square centimetre of the opposite faces of each body, and thus if one body be fixed the other body will be forced away from it until the radiometer force becomes small, when, if there be a restoring force, the body will return to be forced away again by the radiometer force. Thus vibrations depending on the vibrational properties of the body will be set up.

If, as in the present case, gas be evolved, this will give an additional force which may be sufficient for the explanation of the vibrations. I wish only to emphasize that the radiometer force should also be taken into account in the explanation.

Mr C. R. Darling. I should like to ask whether Miss Waller is of opinion that the vibrations are not in any way due to alternate contraction and expansion, as in the case of the Trevelyan rocker. Can invar, which is practically non-expansive, be made to vibrate by touching with a cold solid?

The extension of this work to the investigation of the properties of metals at different temperatures may prove of great service, as the experimental work is simpler than that involved in other methods having this object, and the results are equally certain.

Mr T. Smith. I should like to call attention to a misleading word which enters into Miss Waller's paper by the merest chance. About twenty years ago one of my colleagues at the National Physical Laboratory had occasion in a paper to mention Töpler's Schlieren method. Shortly afterwards the late Lord Rayleigh referred to

* *Z. f. Metallkunde*, 6, 185 (1929).

this, and pointed out that the method should be known as Foucault's. Since that time I have endeavoured, though not with unvarying success, to persuade members of the N.P.L. staff who apply this method in various ways to adopt Lord Rayleigh's advice. It would help towards a correct understanding of the history of physics, which is so difficult nowadays for students to acquire, if we avoided using terms which tend to mislead, of which many examples besides "schlieren" might be mentioned. I suggest that the Physical Society should, as opportunities occur, call attention to widespread errors of this kind, and endeavour to bring about their general correction.

AUTHOR'S reply. I should like to thank Prof. Knudsen for his interesting remarks regarding the magnitude of the radiometer pressure and its possible contribution to the production of vibrations. It is not, however, possible to get any sign of vibration when a heated metal is touched with another cold body which does not emit gas. The mean free path of air being only about 10^{-5} cm. at atmospheric pressure, and the contact being necessarily light and between two small areas, one of which is irregular, it would appear that the radiometer force must be very small.

Mr Darling raises an important point. I have repeated the experiments, and do find that it is easier to produce notes in iron than in invar, though the pure sustained notes are about equal in loudness. The two materials are comparable as regards thermal conductivity, elasticity, density and vibrating properties, and having regard to experiments made on other metals I am inclined to think that a large coefficient of expansion is a contributory though not a principal cause of the production of vibrations.

With regard to Mr T. Smith's remarks I agree that scrupulous care should be taken to see that honour is given where it is due. A. Toepler introduced the term *Schlierenapparat** to describe the apparatus which was used to detect the *Schlieren* or streaks in imperfect glass. He has added footnotes in both publications, from which we learn that Kirchhoff drew his attention to Foucault's work on the examination of optical surfaces published in the *Annales de l'Observatoire Impérial de Paris*, **5**, 203. He recognizes that his and Foucault's methods are very similar in principle. The word Schlieren does not occur in most German dictionaries; its meaning is given as "streaks (in glass and igneous rocks)" in Patterson's *German-English Dictionary for Chemists*. Since therefore the word is not an ordinary one and the method promises to be used commercially, and since Foucault and Toepler's work was independent and devised for different purposes, it would seem better to speak of the Foucault-Toepler method.

* *Ann. d. Phys.* **128**, 126 (1860); **131**, 35 (1867).

ADDENDA TO DISCUSSIONS

Discussion of paper entitled "The influence of atmospheric suspensoids upon the earth's electric field," by H. L. Wright, M.A., *Proc. Phys. Soc.* **45**, 152 (1933).

The Author. The following footnote appears on p. 154: "According to an investigation carried out by Prof. Whytlaw-Gray, on behalf of the Atmospheric Pollution Research Committee of the Department of Scientific and Industrial Research, the number of particles given by the Owens instrument is about a fifth of the number present in the air and large enough to be seen by a microscope. The details of this investigation have not yet been published." This statement, which refers to a comparison of the jet dust-counter with a method of counting, developed in the course of Prof. Whytlaw-Gray's investigation, in which the particles are deposited by thermal action, was inserted as having an obvious bearing on the subject-matter of my paper. I am informed, however, that the result which was quoted was both premature and incorrect. Subsequent work has indicated that the two methods give counts of the same order. It is regretted that the quotation of the result of what, as I now learn, was in fact an incomplete investigation should have given currency to a misleading statement concerning the performance of the jet dust-counter.

Discussion of paper entitled "A note on the Kerr cell," by E. E. Wright, B.Sc., *Proc. Phys. Soc.* **45**, 469 (1933).

Dr A. Narath. In dealing with the distortions obtained in modulating the characteristic of a Kerr cell, the author appears to have been unaware of certain earlier papers dealing with the same subject. For instance F. Fischer and H. Lichte* have ascertained the distortions of a Kerr cell in terms of Fourier components by the use of Bessel functions. This deduction is identical with that of E. E. Wright. In a different way the present writer in the same year calculated the non-linear distortions in a Kerr cell, tested the deduced relations by experiments, and verified them†. The above papers go further than the paper under discussion, since in addition to the percentage of harmonic waves the rectifying action is calculated and the deduced relations have undergone a thorough experimental test and verification.

Author's reply. The method of effecting the Fourier analysis adopted by Fischer and Lichte differs somewhat from that employed by me, and none of the Fourier amplitudes are given explicitly. Further, the analysis is used only to deter-

* *Tonfilm, Aufnahme und Wiedergabe nach dem Klangfilm-Verfahren*, pp. 236–238. (Hirzel, Leipzig, 1931.)

† Lecture at the Deutscher Physikertag, Bad Elster, in 1931; *Z. f. technische Physik*, **13**, 17 (1932).

mine the distortion for the two bias values, $V/2$ and $V/\sqrt{2}$, and not for determining he position of minimum distortion. Apart from these differences, however, I have o admit that this reference is very pertinent to my own work. On the other hand, I would like to point out that the book is not readily available in England; for example, here is no copy in the Library of the British Museum. Dr Narath's own paper is very extensive and interesting but I consider that it is less relevant as an anticipation of my analysis, since the expressions obtained in it for the amplitudes of the components are in the form of doubly infinite series and are, I think, less suited to computation than mine, in spite of the fact that the use of Bessel functions is avoided.

REVIEWS OF BOOKS

A Text-book of Physics. Volume 2, *Heat and Sound*, by E. GRIMSEHL. Edited by R. TOMASCHEK. Authorized translation from the 7th German edition by L. A. WOODWARD. Pp. xii + 312 with 225 figures, numerical tables and examples. (London: Blackie and Son, Ltd.) 12*s.* 6*d.* net.

This is the second of a comprehensive series of text-books of physics of which the first three have now appeared. References are given in Volume 3 (Electricity) to articles in Volume 5 which has not yet been published.

The present volume could be read with interest by apprentice engineers, but the same criticism applies that has been made with regard to Volumes 1 and 3: the presentation is not sufficiently formal for serious students of physics. For them many of the otherwise excellent figures will be wasted, as for example that of the Babcock and Wilcox boiler with superheater on p. 157, and the force-transmission to the crank of a six-cylindered motor, p. 171, which is incomprehensible. This engineering aspect is marked in table IV *b*, p. 291, which gives the temperatures at which the vapour pressure of water is 1, 2, 3, etc. kg./cm.^{2}

The subject-matter of heat goes as far as elementary considerations of entropy, but the fact that the simple Carnot cycle is nowhere represented in ϕ, $t°$ coordinates, that the fundamental researches of Rowland and of Reynolds and Moorby and their significance in the philosophy of the subject are absent, and also that the work of Callendar and Barnes is relegated to an appendix, will be sufficient to indicate that the publishers' claim that the standard is that of a degree in physics is scarcely justified.

There are some excellent figures and diagrams in the chapters on wave motion and sound. The structure of the text calls for little comment. The physical presentation is better here, but there is a marked absence of any references to the mass of very interesting modern researches.

The index needs revision. There is no reference to the singing flame (p. 296) or to viscosity (p. 291). L. S.

A Text-book of Physics. Volume 3, *Electricity and Magnetism*, by E. GRIMSEHL. Edited by R. TOMASCHEK. Authorized translation from the 7th German edition by L. A. WOODWARD. Pp. xiv + 660 + pp. 10 of numerical examples. (London: Blackie and Son, Ltd.) 25*s.* net.

This book, like the volume on mechanics by the same author, should appeal to a much wider circle than those working for "a degree in physics," to quote from the publishers' advertisement. One can browse amongst the pages, each of which has its interesting pictures or diagrams, and some pages have three or four diagrams. These range from a full-page catalogue reproduction of a modern screening couch and X-ray apparatus down to a picture of a tumler switch.

The volt-ampère system of units is used throughout. As the author says in the preface, every physicist employs only instruments calibrated according to this system and has an intuitive appreciation of the significance of the quantities employed. Thus we see on p. 178 the statement "The unit of magnetic field strength H is that field strength which exists in a long solenoid through which is passing a current of one ampère per cm. length." I cannot gather the meaning of this. The system of units throughout the book has this

basis. The only parallel case that has come into common usage in physics recently is the nomenclature for the velocity of an electron in terms of electron-volts, but I doubt if the accepted usefulness of this term would be taken to be a justification for altering the whole fabric of systematic physics.

In spite of the fact that this book presents a very interesting appearance, the careful students of physics reading it will be on nodding but not familiar terms with principles. In the long run it will be far better to jettison all the repetitive detail that does not help to advance the subject. L. S.

The Development of Physical Thought, by L. B. Loeb and A. S. Adams. Pp. 648. (London: Chapman and Hall.) 23*s*.

A welcome may be given to this general survey of physics, which formed the basis of lectures given to a class "open to all letters and science students regardless of class and irrespective of previous exposure to high school physics." The book is not exhaustive, but is divided into six chapters, of which the first is historical, giving an account of some of the principal movements of scientific thought from early times to the foundation of the Royal Society and after.

On p. 50 is a graph which is supposed to give the number of significant discoveries as a function of time in years. Whether this curve has much validity is to be doubted, but it does illustrate the enormous growth of research towards 1900. It is then asked if this acceleration of output is to continue, and the conclusion is suggested that knowledge will increase at an almost constant rate.

It is questionable whether "in Oxford even now, a scientific worker or student is looked upon with scorn by the classicists."

Chapter 2 is devoted to mechanics. Admirable features, in contrast to the general development of the subject, are the detailed treatments of the inclined plane and lever to which many measurements, systematically arranged, are appended.

On pp. 157 and 445 an account is given of the Michelson-Morley and kindred experiments, such that the reader is left with the impression that no ether wind has been detected. Later work by Miller, Courvoisier, etc. would indicate that the question is perhaps not finally settled. On p. 165 it is stated that "it was not until 1919 that President Campbell of the University of California and members of his expedition were able to make a successful and accurate set of observations of the positions of the stars in the neighbourhood of the sun during an eclipse." Some reference to the work of Dyson, Eddington and Davidson seems to be called for.

Chapters 3 and 4 deal with heat and classical electricity respectively. Chapter 5 is given to light, which is rather briefly treated in some thirty pages. The spectroscope, for example, has a very scant description. Chapter 6 gives a very full and interesting account of modern electrical work on the structure of matter and allied topics. Many photographs illustrating the work of C. T. R. Wilson, Blackett, Feather and others are reproduced from the pages of the *Proceedings of the Royal Society*. Considerable attention is devoted to the atomic discoveries of the last few years, and it would be a dull student indeed, who could not find an interest in this last almost romantic chapter.

The book contains over one hundred figures, those of apparatus being simplified to the last degree.

Finally there is a bibliography of literature suitable for further reading, a table of physical constants and nearly thirty pages of index.

To supplement work with a more conventional text-book, this volume giving sidelights on physics as a development of thought will be extremely attractive to any student with an enquiring mind. J. E. C.

Analytic and Vector Mechanics, by HIRAM W. EDWARDS, Ph.D. Pp. x + 428. (McGraw-Hill Book Company, Inc., 1933.) 24*s*. net.

This book gives a general account of the principles and methods of dynamics up to a standard which is rather higher than that of the general B.Sc. degree in the University of London, but not quite sufficient for the Special Honours degree. There are many points in the book to be greatly welcomed. The introduction of brief historical references does a great deal to add a certain human interest to a subject which can be rather arid and abstract, and the quotation from Newton on p. 128 will come as a surprise and an inspiration to many students.

Again, the introduction of vectorial methods is extremely useful in abbreviating some of the demonstrations, particularly in obtaining the fundamental equations of uniplanar rigid kinetics. It seems to the reviewer, however, that the use of vectors does very little to simplify the theory of the general motion of a rigid body as exemplified in tops and gyroscopes; and that the deduction of Kepler's first law by vectorial methods given on pp. 262 and 263 is unnecessarily clumsy. The method employed in the book makes use of the concept of the reciprocal of a vector, but this concept is nowhere defined, nor are the properties of the reciprocal of a vector investigated in any other part of the book. This is particularly unfortunate in view of the fact that the method employed in these pages is only successful because the vector denoted by A^{-1} is perpendicular to the vector r.

The equations of motion are treated on the basis of Newton's laws made palatable by the method used by Appell. No critical account of the foundation of the subject is attempted, and the reader will be bitterly disappointed with the paragraph headed "Standard Reference System" on p. 2.

D'Alembert's principle is not explicitly mentioned until the last chapter, although it is of course used in obtaining the equations of motion of a rotating body in chapter 8. In fact the proof of D'Alembert's principle is left as an exercise for the student!

The greater part of this work is taken up with dynamics, but there are also included one chapter each on statics, attractions and potential, and vector fields. The reader accustomed to the treatises of Loney and Lamb will be disappointed at the small number of exercises included in the book. However, he will also be pleased to note a few original examples, e.g. the deflection of an alpha particle by a movable nucleus, the sky rocket, the water stream and bucket.

Personally I feel that this book falls between two stools. There is not sufficient analysis for the mathematician and there is not sufficient experimental illustration for the physicist. Finally, I have very serious doubts whether the introduction of vectors at this stage does effect any real simplification. In spite of all these criticisms, it must of course be recognized that the author had a very difficult task before him. None of the existing text-books on this subject is completely satisfactory, and Prof. Edwards has made a most courageous attempt to provide a more acceptable text-book on mechanics. G. T.

Practical Acoustics for the Constructor, by C. W. GLOVER. Pp. xi + 468. (London: Chapman and Hall.) 25*s*. net.

The past few years have seen published a spate of books upon the subject of architectural acoustics; their outlook, their treatment, their length and their style have varied considerably but among them the present volume must be regarded as unique. In character it is essentially a book of applied science and contains those minutiae of detail and that wealth of illustration which are associated with a volume devoted to engineering design.

In his preface the author professes the aim of providing a vade mecum for the practical man and, bearing this in mind, we must condone the absence of formal proofs, which the purist might regard as a sin of omission, and also consider leniently sins of commission in the nature of diagrams popular rather than scientific in their appeal. It must not, how-

ever, be imagined that the book belongs to the "Wonders of Science for the Man in the Street" series, as Mr Glover is equally at ease in constructing the exercises appended to each chapter, and in summarizing in a most valuable manner the Bureau of Standards sound insulation values for various partitions, as he is in indicating the method of concealing nail heads when fixing board absorbents.

Adhering to convention, the opening chapters are devoted to a general discussion on the subject of sound. While we note a measure of confusion between intensity and loudness, and some numerical inexactitude associated with the description of stationary waves, we heartily commend, as a device appreciated by students, the use of a diagram of the genealogical-table variety which sets out the scope of the subject in a clear and instructive manner. Broadly speaking the matter reported in the body of the book is that common to other treatises on the same subject, except that we do not remember having encountered elsewhere chapters devoted to the organ and its position in an auditorium, or to aircraft noise: we also consider that the photographic illustrations and drawings of constructional details are more prolific than usual.

Not the least remarkable features of the volume are the chapter on acoustical specification clauses and the series of appendices, which, to the practitioner, must be worth almost the cost of the book. The first appendix is a somewhat haphazard list of books and papers concerned with acoustics, vibration and noise; the second a summary of acoustical defects together with their remedies; the third a list of no fewer than seven hundred and two absorbent materials together with their absorption coefficients over considerable frequency ranges; and the last, illustrated particulars of over forty acoustical materials. Our only possible grievance is that the author, for fear of controversy and invidious comparison, was deterred from inserting cost data in appendix 3, but nevertheless we must conclude by expressing to Mr Glover our appreciation of his generosity in divulging so many mysteries of his craft in such a painstaking and interesting manner. E. J. I.

Gyroscopic Stabilisation of Land Vehicles, by J. F. S. Ross, B.Sc., Ph.D. Pp. vii + 172. (London: Edward Arnold and Co.) 14*s*. net.

This book discusses in the first 100 pages the mathematical requirements for the stability of such vehicles. From the equations of motion the usual periodic equation, a linear differential equation of the fourth order, is derived. The conditions that this equation shall represent two damped oscillations are then discussed at unnecessary length*. Numerical examples are given and suggestions for a design to incorporate the essential controls for stability are added.

The rest of the book is devoted to a criticism of the designs of the pioneers in this field of work, Brennan, Scherl, Schilovsky and others. No description of the apparatus used is given, but the patent descriptions are discussed in the light of the essential requirements shown by the analysis of the periodic equation. The author evidently does not like discontinuous control. Brennan's ingenious reversal of the friction couple by means of his dead and live rollers and Schilovsky's spasmodic engagement of the precessional torque are disapproved.

If ever the monorail is adopted in practice, the method of controlling the gyroscopic precession may be continuous or discontinuous, but the reasons retarding its adoption are not due to any failure to understand the scientific requirements. The parallel problem of the damping of the rolling of ships is an engineering success.

The book will appeal mostly to students of a mathematical bent who will find many interesting problems in it. An excellent glossary is given at the end.

* This equation occurs in many gyroscopic problems and is dealt with very shortly in a paper by Prof. Sir J. B. Henderson in the *Proceedings of the International Mathematical Congress at Toronto*, 1924.

Elements of Engineering Acoustics, by L. E. C. Hughes. Pp. xi + 159. (London: Ernest Benn, Ltd.) 8*s*. 6*d*. net.

The scope of this book is indicated by its seven chapter-headings, which are: Sounds and Sound Systems, Sound-Field and Sound Measurements, Reproduction of Sound, Electro-Acoustic Measurements, Microphones, Amplifiers, and Reproducers; and its substance has already been published by the author in a series of articles in the Students' Section of the *Electrician*. On account of this origin a somewhat disjointed text might be anticipated, but in actual fact Dr Hughes, in his rôle of editor, has given us a general if slightly dogmatic discussion of his subject in the form of a continuous narrative, which, regarding perspicuity of text as the hall mark of a good book, we should have preferred to find more frequently broken by paragraph headings.

We commend in particular the final chapters dealing with the microphone-amplifier-reproducer system, and note with interest what the author has to say on the subject of *acoustic distortion*.

While we may perhaps be permitted to deprecate mildly the inclusion of a picture of (mainly) four valves bearing the legend "Loaded Push-Pull Power Amplifier," and the designation of ultrasonic waves of high intensity as a "form of death ray", our repudiation must take a stronger form when we learn that to obtain different frequencies we merely require suitable taps on an a.-c. mains transformer. On the other hand we receive with personal satisfaction Dr Hughes's somewhat unorthodox contention that musical experts or musicians are of no use in assessing the quality of reproduction of music save when it is very bad.

If we may appear hypercritical it is because we welcome a book covering the field treated by an author who speaks with authority and whose message we would, for others as well as ourselves, desire to be as easily assimilable as possible. E. J. I.

Recent Advances in Physical Chemistry, by S. Glasstone, Ph.D., D.Sc., F.I.C. Pp. viii + 498. (Second Edition. London: J. and A. Churchill.) 15*s*.

The fact that a new edition of Dr Glasstone's excellent summary of some of the more notable advances in physical chemistry should be called for within two years of the publication of the first edition is sufficiently convincing testimony of its value. The author has taken the opportunity to include in the new edition a large mass of fresh material, and has discussed briefly the application of wave mechanics to problems of valency and to the calculation of the energy of activation; nuclear disintegration, the neutron and the positive electron; the influence of free and restricted rotation on dipole moments; molecular beams; potential-energy curves; atomic reactions; the kinetics of photochemical reactions; activated and discontinuous adsorption; surface potentials; and the mobility of surface molecules. A very useful chapter on solubility has disappeared, but even so the present edition is longer by some thirty pages than its predecessor. Physicists who desire to know something of the central problems of modern physical chemistry and of the outlook of the physical chemist of to-day will find Dr Glasstone's book of great service. A. F.

An Introduction to Thermodynamics for Chemists, by D. Johnston Martin. Pp. vii + 343. (London: Edward Arnold and Co.) 16*s*. net.

We may not all agree that it is a good thing to write books on thermodynamics for engineers, for chemists, for horticulturists..., but it is certain that the chemist has a compelling interest in many topics that are best treated with the assistance of a thermodynamical argument. A book, therefore, which gives a perfectly sound and general discussion of the bases of modern thermodynamics, and afterwards proceeds to apply these

principles to physico-chemical topics, has ample *raison d'être*. The author's introductory chapters are scholarly, and he is thoroughly alive to the presence of those pitfalls which lurk ready to entrap the unwary and facile reasoner. In the part of his text which deals with applications he has treated very fully dilute solutions, the activity method of treatment of solutions, the theories of strong electrolytes, chemical affinity, and heterogeneous systems.

Following the work of Lewis and Randall, the author has discussed very fully the distinction between free energy and maximum work. He has produced a very serviceable volume, which should find wide acceptance. A. F.

Introduction to Physical Chemistry, by ALEXANDER FINDLAY. Pp. vii + 492. (London: Longmans, Green and Co., Ltd.) 7*s*. 6*d*.

Prof. Findlay is not afraid to begin at the beginning, and has given us, in rather less than five hundred pages, a text-book which is a model of what such a book should be. Starting with the very elements of the subject, he develops a course which holds the balance admirably between the old and the new, between theory and experiment, and leaves the student with nothing to unlearn when he proceeds to more specialized study. The topics of a course of degree standard are treated of succinctly and thoroughly, very full references to original sources are scattered through the text, and an appendix furnishes a number of numerical examples which may serve to test the reader's knowledge. The book may be unreservedly recommended. It is well produced and illustrated and is, at the price, remarkably good value. A. F.

Correction Tables for Use with Platinum Resistance Thermometers, by G. S. CALLENDAR and F. E. HOARE. (London: Edward Arnold and Co.) 1*s*.

These tables have been computed so that the labour of reducing readings with a platinum resistance thermometer is reduced to a minimum. Nowadays when a platinum resistance thermometer is selected for the measurement of moderately high temperatures, the reason is that by means of it temperature-measurements of high precision can be made. On this account it is desirable to have tables for the conversion of platinum temperatures to the international scale worked out to a high order of accuracy, and in the present tables the corrections are given to several places of decimals. The main table covers the range 0 to 1000° C. In addition a table is given of the corrections to be applied when in the well-known difference formula the delta is not exactly 1·5.

To anyone concerned with resistance thermometry the table will prove a sound investment. E. G.

Photograms of the Year 1933, edited by F. MORTIMER, Hon. F.R.P.S. Pp. 24 with 64 plates. (London: Iliffe and Sons, Ltd.)

"This here science, it keeps going on," and it is odd to think, in an age of cinematographic cameras capable of taking records at the rate of thousands a second, that in the youth of our parents a visit to the photographer held almost as many terrors as a twentieth-century visit to the dentist. The maiden of the 'fifties or the 'sixties, swathed in voluminous and unsightly garments, and clamped, as to the back of her head, in a steel contraption designed to hold her steady during an exposure, has left us photographic records of herself which are studies not so much of a human being, as of a "head sticking out of a dress."

That is an accusation which certainly cannot be laid against many of the delightful personal studies—sometimes graceful, sometimes bizarre—in the volume under review. Seascape and landscape; the human model; still life studies, orthodox and unorthodox;

the suggestion of immense force latent in a crank-shaft and locomotive driving-wheel, and of skill and artistry in a hand engaged in fine lace manufacture—nothing comes amiss to the modern camera artist. To the lore of the man with the palette he has added a technique peculiarly his own, which is abundantly exemplified in the reproductions given in this most interesting and valuable record of the progress of camera art in 1933. A. F.

Liquid Crystals and Anisotropic Melts. A General Discussion held by the Faraday Society. Pp. iv × 204. (London: Gurney and Jackson, 1933.) 12*s.* 6*d.* net.

The present volume, reprinted from the *Transactions of the Faraday Society*, contains twenty-four papers and a general discussion. The continental authors include Vorländer, Kast, Ornstein, Ostwald, Rinne and Zocher, and the English authors Bernal, W. H. Bragg, Crowfoot, Fowler, Lawrence and Malkin. Ornstein and Zocher discuss rival theories of liquid crystals, Rinne emphasizes their biological importance, Lawrence deals with lyotropic mesomorphs and Bernal and Crowfoot contribute a study of the true crystalline phases of substances which form liquid crystals. In their paper, Bernal and Fowler deal with the structure of ice. The accepted structure has been modified, the unit cell being much larger and the structure molecular; four molecules are arranged in approximately tetrahedral fashion around a fifth. This arrangement in conjunction with Prins's theory is used to explain the X-ray diffraction patterns of water and their change with temperature. Water changes from something approximately ice-like at low temperatures to a quartz-like structure at ordinary temperatures, and further heating tends to produce close packing. The whole volume will be of considerable interest to those concerned with structural physics and chemistry, particularly from the point of view of the transition from the solid to the liquid state. J. R. P.